

=* fil wpix

FILE 'WPIX' ENTERED AT 14:26:27 ON 12 FEB 2001
COPYRIGHT (C) 2001 DERWENT INFORMATION LTD

FILE LAST UPDATED: (7 FEB 2001) <20010207/UP>

>>>UPDATE WEEKS:

MOST RECENT DERWENT WEEK: .000108 <200108/DW>

DERWENT WEEK FOR CHEMICAL CODING: .000108

DERWENT WEEK FOR POLYMER INDEXING: .000108

DERWENT WORLD PATENTS INDEX SUBSCRIPTION FILE, COVERS 1963 TO DATE

>>> FOR UP-TO-DATE INFORMATION ABOUT THE DERWENT CHEMISTRY RESOURCE, PLEASE VISIT

[<<<](http://www.derwent.com/chemistry/resource/index.html)

>>> FOR DETAILS OF THE PATENTS COVERED IN CURRENT UPDATES, SEE [<<<](http://www.derwent.com/covcodes.html)

=* d fili all abeq tech tot

LIC3 ANSWER 1 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN WO01-071346 (06) WPIX

DNIN WO01-053953 DMC C1001-019999

TI Identifying biological macromolecules, e.g. **proteins**, from mass data, by determining frequency distribution of t scores representing similarity of data for known and unknown compounds.

DC 504 016 **T01**

IN SHAIT, B T; ERIKSSON, J; BENNOE, B

PA (UYEQ) UNIV ROCKEFELLER

CYC 51

PI WO 2000077712 A1 20001221 (200108)* EN Esp G06F019-00 <--
FW: AT BE CH CY DE DK ES FI FR GB GF IE IT LU MC NL PT SE
W: CA JP US

ADT WO 2000077712 A1 WO 2000-US16633 20000615

PRAI US 1999-333726 19990615

IC ICM **G06F019-00**

AB WO 200077712 A UPAB: 20010207

NOVELTY - Generating a frequency distribution (FD) of scores (S), related to random identifications of biological molecules (I), for a particular experimental condition (EC), is new.

DETAILED DESCRIPTION - Generating a frequency distribution (FD) of scores (S), related to random identifications of biological molecules (I), for a particular experimental condition (EC), is new. The method comprises:

(a) generating mass data at EC, for known (I) in a database, and for a hypothetical (I);

(b) comparing the data to calculate S that is a **function** of the similarity of the data;

(c) selecting S that represent a high degree of similarity;

(d) repeating the method for other hypothetical (I); and

(e) determining the frequency of selecting each S to generate FD.

INDEPENDENT CLAIMS are also included for the following:

(1) the novel method in which the hypothetical (I) in the first step is replaced by a (I) randomly selected from the database;

(2) the novel method for identifying, at a particular significance level, an unknown (I);

(3) determining statistical significance of an identified score for (I);

(4) FD prepared by the novel method;

(5) identifying a (I) at a particular EC and significance level, comprising:

(a) selecting a significance level that represents a level of confidence in a biological molecule identification;

- (b) cleaving the unknown biological molecule into constituent parts;
- (c) generating mass data for these constituent parts;
- (d) comparing the mass data with mass data generated for the EC from known biological molecules of a database;
- (e) calculating scores for each comparison, which are a **function** of similarity between mass data of the unknown and known molecules;
- (f) selecting a score generated in (e) which corresponds to a comparison which denotes a high degree of similarity, and which corresponds to a particular known molecule;
- (g) comparing the score selected in (f) with FD of scores for EC, the distribution relates to random biological molecule identifications, and has a critical score associated with it which corresponds to the significance level; and
- (h) determining if the score in (f) is equal to or higher than the critical score;
- (i) determining the statistical significance of an identified score for (I), comprising:
 - (a) selecting a significance level that represents a level of confidence in a biological molecule identification;
 - (b) calculating a score associated with an unknown biological molecule, the score is a **function** of similarity between mass data of the unknown and known molecules;
 - (c) comparing the score with a score FD, the distribution is generated by comparing mass data of hypothetical biological molecules with mass data generated for known molecules, and the FD has a significance level associated with it; and
 - (d) determining if the score associated with the unknown molecule identification is within the significance level; and
- (j) a computer program system for performing the novel method.

USE - The method is used to identify unknown (I), particularly **proteins**, nucleic acids and polysaccharides, from mass spectroscopy data and to assess the reliability of the identification.

Dwg.0/14

FS CPI EPI
 FA AB; DGN
 MC CPI: B04-C02; B04-E01; **B04-N04**; B11-C08; B11-C08A;
B12-K04E; D05-H09
 EPI: T01-J

TECH UPTX: 20010207

TECHNOLOGY FOCUS - INSTRUMENTATION AND TESTING - Preferred Process: FD is particularly a probability density **function** and mass data are generated by computer or, for hypothetical (I), by mass spectrometry. The method is used to generate 1-1010, particularly 100-107 S values. EC defines mass data results from chemical degradation of (I) (particularly enzymatic degradation with trypsin), including the effectiveness of the degradation, or a particular accuracy for mass determination. The comparison of data is limited to known (I) within selected ranges of either mass or isoelectric point, especially to known (I) that are modified, e.g. for **proteins** post-translationally modified. To generate hypothetical (I), at least one known (I) is selected from the database (preferably in the mass range up to 3 MD), mass data generated and those masses selected that correspond to at least one constituent part of the known (I). The procedure is repeated for different (I) until sufficient masses (particularly 1-1000) have been selected to allow generation of hypothetical (I) which consists of a set of constituent parts that is different from all sets of constituents in known (I). Selection of (I) and of the constituent parts is random. Fragment mass data may also be collected for at least one constituent part of known and hypothetical (I), and in this case the comparison is between fragment mass data and EC then defines the energy used to generate these data. In the method of (2), a significance level is chosen that represents a level of confidence in identification of (I), then unknown (I) cleaved into its constituent parts (especially in a predictable manner) and mass data generated for these parts. These data are then compared with data from known (I) to generate an S which is compared with an FD of scores, at EC,

that is related to random identification of (I), and associated with a critical score corresponding to the significance level. The unknown (I) may be a mixture. The fragment mass data are acquired by vibrational or electronic excitation, specifically collisions with electrons, photons, gas molecules or a surface. In the method of (3), S from an unknown (I) is compared with an FP, associated with a predetermined significance level and generated by comparing mass data from known and hypothetical (I). Any S within the significance level (which particularly denotes a probability that identification of (I) is false) is determined.

Preferred Materials: (I) are **proteins**, nucleic acids or polysaccharides.

L123 ANSWER 2 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 2001-007103 [01] WPIX
 DNN N2001-005104 INC 2001-001769
 TI Biological marker and phenotype identification system, useful in drug development, using processor to correlate data regarding e.g. cell populations, soluble factor levels and clinical parameters.
 DC B64 B16 J04 S03 T01
 IN ALLISON, A; BRUNKE, K J; DIETZ, L J; KANTOR, A B; NATAN, M J; RINGOLD, G
 PA (SURF-II) SURROMED INC
 CYC 89
 PT WO 200065472 A1 20001102 (200101) EN 194p G06F017-00 <--
 RW: AT BE CH CY DE DK EA ES FI FR GB GR GM GR IE IT KE LS LU MT MW NL
 OA PT SD SE SL SI US SW
 W: AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE DK DM EE ES
 FI GB GE GH GM HR HU ID IL IN IS JP KE KS KP KR KZ LC LK LR LS
 LT LU LV MA MD MM MN MW MA NO NZ PL PT RU SU SD SE SG SI SK SL
 TJ TM TR TT TC UA WG US VI VI SA SW
 ADT W. 200065472 A1 WO 2000-0311296 2000426
 PFAI US 2000-175075 20000107; US 1999-131195 13330126
 IC ICM G06F017-00
 AB WO 200065472 A UPAB: 20001-30
 NOVELTY - A biological marker (BM) identification system comprising an integrated database containing several data categories and data from various organisms corresponding to the data categories, is new.
 DETAILED DESCRIPTION - A biological marker (BM) identification system comprises:
 (a) an integrated database containing several data categories, i.e. levels of cell populations, cell associated molecules and/or soluble factors in a biological fluid and information associated with clinical parameters of an organism; and
 (b) data from various organisms corresponding to the data categories.
 A processor correlates data within the categories, to identify the category (or categories) indicating normal biological or pathogenic processes or responses to drug intervention; this category (or these categories) is/are BM.
 INDEPENDENT CLAIMS are also included for the following:
 (1) a method for identifying a BM for a given disease or medical condition (GD/MC), comprising correlating information associated with several categories (as in (a) above) from several organisms, at least some of which have GD/MC and identifying a data category (i.e. BM) by which the presence of GD/MC can be identified;
 (2) a phenotype of an organism comprising several biological parameters, i.e. the results of at least 10 (preferably at least 40) assays relating to cell populations and/or cell associated molecules, the results of at least 20 (preferably at least 40) assays relating to soluble factors and clinical parameters;
 (3) a phenotype of a class or subclass of organisms, comprising parameters as in (2) for each member;
 (4) a system for creating the phenotype of an organism, involving obtaining parameters as in (2);
 (5) a method for evaluating the effect of a perturbation on an organism (or on a class or subclass of organisms), involving comparing the information in the phenotype, as in (2) or (3), of the organism(s) before and after the perturbation;

- (6) a system for the identification of BM's of a GD/MC in an animal model, involving an integrated database and processor as above;
- (7) a method for identifying a BM for GD/MC in a human, involving determining if a BM in an animal model as in (6) is diagnostic or prognostic of GD/MC in a human;
- (8) a method for assaying a candidate drug, involving treating an animal model with the drug and evaluating the effect on a BM as in (6);
- (9) a method for monitoring the results of a clinical study in humans with a GD/MC, involving identifying BM's in a human which are homologs of BM's identified in animal models of GD/MC;
- (10) a method for designing an improved animal model for a GD/MC in humans, involving identifying human BM's relative to the GD/MC and tailoring the animal model to simulate GM/MC more accurately by elevating or reducing the levels of animal homologs of the human BM;
- (11) a method for identifying an animal model of a GD/MC, involving comparing phenotypes (as in (2)) for potential animal models and an organism having the GD/MC, to identify the most closely simulating animal model phenotype; and
- (12) a method for evaluating the effects of the effects of a genetic alteration in a plant or animal, involving comparing information in phenotypes (as in (2)) for the genetically altered and non-altered organism to identify changed parameters.

USE - The systems and phenotypes are useful in drug development. The diseases involved are specifically asthma, allergy, multiple sclerosis or especially rheumatoid arthritis (all claimed). More generally the phenotypes may be of humans, animals, plants or viruses (all claimed); they may also be used for evaluating the effects of a genetic alteration on a plant or animal.

ADVANTAGE - The technology is supplied for providing quantitative, sensitive, reproducible and rapid measurements of multiple and diverse parameters which can accurately profile an organism's phenotype or a patient's disease status and response to therapy. More cost-effective drug development is possible. The biological parameters can be identified from small samples of blood.

DESCRIPTION OF DRAWING(S) - The figure is a schematic representation of the types of information assimilated to obtain a biological marker identification system.

Dwg.1/9

FS CPI EPI
 FA AB; GI; DCN
 MC CPI: B04-F01; B04-F01; B11-C08; B12-K04; D05-H09; J04-B01
 EPI: S03-E14H9; T01-J

TECH UPTK: 20001230
 TECHNOLOGY FOCUS - INSTRUMENTATION AND TESTING - Preferred System: The data for levels of cell populations and/or cell associated molecules are obtained by microvolume laser scanning cytometry. The data for levels of soluble factors (preferably **proteins** or small molecules) are obtained by microvolume laser scanning cytometry, immunoassays or mass spectrometry. The data categories may also include genotype information. The information associated with clinical parameters is selected from age, sex, weight, height, body type, medical history, environmental factors and categorization of GD/MC.

L123 ANSWER 3 OF 73 WPIX COPYRIGHT 2001 DEFVENT INFORMATION LTD
 AN 2000-686352 (67) WPIX
 DNN N2000-507949 DNC C2000-208396
 TI Analyzing molecule and **protein** diversity using a computer method comprising defining a set of constraints on possible target surfaces and defining a set of all theoretical target surfaces.
 DC B04 D16 S03 S05 T01
 IN MOALLEMI, C C; WINTNER, E A
 PA (NEOGEN-N) NEUROGENESIS INC
 CYC 91
 PI WO 2000060507 A2 20001012 (200067)* EN 103p G06F017-50
 FW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL
 OA PT SE SL SE TS UG SW

W: AE AG AL AM AT AU AZ BA PB BG BP BY CA CH CN CR CU CS DE DK DM DZ
 EE ES FI GB GE GH GM HR HU IL IL IN IS JP KE KG KP KR KZ LC LK
 LR LS LT LV MA MD MG MY MN MW MX NG NZ PL PT RU SD VE SG SI
 SK SL TJ TM TR TT TZ WA UG US US VN YE ZA SW

AU 2000044511 A 1006102A (2006102)
 G06F017-50 --
 ADT WO 2006060507 A2 WO 2006-060507, 0000331; AU 2000044511 A AU 1006-44511
 20000331

FDT AU 2006060507 A Based on WO 2006060507
 PRAI US 1094-107486 1994-04-01

IC ICM G06F017-50
 AB WO 2006060507 A UPA8: 2006102A

NOVELTY - A computer-based method comprising defining a set of constraints on possible target surfaces, and defining a fully enumerated set of theoretical target surfaces under the defined constraints, so that each surface has a defined, continuous volume and a defined, continuous surface area, is new.

DETAILED DESCRIPTION - A computer-based method comprising defining a set of constraints on possible target surfaces, and defining a fully enumerated set of theoretical target surfaces under the defined constraints, so that each surface has a defined, continuous volume and a defined, continuous surface area, is new. The method further comprises mapping sets of objects to the fully enumerated set of theoretical target surfaces to define corresponding subsets of the fully enumerated set of theoretical target surfaces, and analyzing an aspect of diversity of the objects based on degrees of similarities and differences among the corresponding subsets.

An INDEPENDENT CLAIM is also included for a computer programmed to determine the chemical similarity of different molecules, the program comprising:

(a) approximating the surface shape of molecules of interest by linking a series of cubes, each having a dimension R, the locations of the cubes being determined by the calculated electron probability density of the individual molecule of interest, each cube sharing at least one of its six faces with another cube, so that there is a specific number of linked cubes which varies for each molecule of interest;

(b) approximating the chemical reactivity of each individual molecule of interest by assigning each cube of each individual molecule of interest, no more than one **functionality** value from M different chemical **functionality** values;

(c) approximating the surface shape and chemical reactivity of a chemically active surface having a volume equal to V by subtracting a number V/R³ cubes of dimension R from a surface, where each cube space shares at least one face with another cube space and where N cube spaces have one of M different chemical **functionality** values;

(d) calculating an attraction value K for each molecule of interest to the chemically active surface; and

(e) calculating a list of overall attraction values to the chemically active surface.

USE - For analyzing molecule and **protein** diversity.

Dwg.0/25

FS CPI EPI
 FA AB; DCN
 MC CPI: B04-N04; B11-C08; B12-K04E; D05-H09
 EPI: S03-E14H5; S03-S02; T01-J06A; T01-J07B
 TECH UPTM: 2006102A

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preferred Materials: The target surfaces comprise negative space target surfaces. The objects comprise positive space object surfaces associated with different molecules. The objects comprise positive space object surfaces associated with different molecules. The objects are mapped by defining corresponding subsets of the set of negative space theoretical target surfaces to which the positive space object surfaces of conformations of molecules are complementary. The aspect of diversity that is analyzed is the difference or similarity between the molecules, which map to those negative space theoretical target surfaces. The objects comprise negative space object surfaces associated with different **proteins**. The objects are mapped by

defining corresponding subsets of the fully enumerated set of negative space theoretical target surfaces to which the negative space object surfaces of the **protein** pockets are similar. The aspect of diversity that is analyzed is the difference or similarity between **protein** pockets, which map to those negative space theoretical target surfaces. Alternatively, the objects comprise positive space object surfaces associated with different molecules and negative space object surfaces associated with different **proteins**. The objects comprise positive space object surfaces associated with different molecules and negative space object surfaces associated with different **proteins**. In the case of molecules, the objects are mapped by defining corresponding subsets of the fully enumerated set of negative space theoretical target surfaces to which the positive space object surfaces of conformations of molecules are complementary. In the case of **proteins**, the objects are mapped by defining corresponding subsets of the fully enumerated set of negative space theoretical target surfaces to which negative space object surfaces of **protein** pockets are similar. The theoretical target surfaces comprise polyhedrons. The objects comprise polyhedrons. The polyhedrons comprise cubes, which are all of the same size and shape. The set of all theoretical target surfaces defines a diversity space within which the diversity of objects can be measured by mapping those objects to the diversity space.

Preferred Methods: The method also includes identifying regions of the diversity space to which no objects map. The method also includes designing molecules that occupy at least one of the unfilled theoretical target surfaces of the diversity space. Complementarity is associated with binding affinities of positive space object surfaces of conformations of molecules to negative space theoretical target surfaces. The constraints comprise volume. The constraints comprise associating each site of the target surface with a pre-selected molecular property. Each of the pre-selected molecular properties is drawn from a larger set of possible molecular properties, e.g. hydrophobic, polarizable, H-bond acceptor, H-bond donor, H-bond donor/acceptor, potentially positively charged or potentially negatively charged. Not all of the sites of the target surface are each associated with a different one of the molecular properties and all of the other sites of the target surface are associated with a common molecular property. The common molecular property is slight hydrophobicity. The degrees of similarities or differences comprise **functional** properties associated with the corresponding subsets of the fully enumerated set of theoretical target surfaces. The degrees of similarities or differences comprise shape properties associated with the corresponding subsets of the fully enumerated set of theoretical target surfaces. The method further comprises defining each of the objects by quantifying molecules into polyhedrons.

L123 ANSWER 4 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 2000-665269 (64) WPIX
 DNN N2000-493033 DNE C1000-201590
 TI Identifying novel nucleic acid molecules encoding **proteins** of interest, and natural language processing and extraction of relational information associated with genes and **proteins** found in journal articles.
 DC B04 D16 J03 T01
 IN FRIEDMAN, C; KALACHIKOV, S; KRAUTHAMMER, M; PICHETSKY, A
 PA (NYC) UNIV COLUMBIA NEW YORK
 CYC 51
 PI WO 2000043687 A1 20001026 (200064)* EN 374p 9018831-00
 FW: AT BE CH CY DE DK EA ES FI FR GB GH GM GP IE IT KE IS LU MC MW NL
 OA PT SD SE SI SG TO US CW
 W: AE AG AL AM AI AU AD BA BB BG BR BY CA CH CN DE CU CZ DE DK EM DZ
 EE ES FI GB GE GE GH GM HF HG IL IL IN IS JP PE KG PP KR KZ LC LK
 LF LS LT LU LV MA MD MG MW MN MW MX NJ NC PL PT RO RU SD SE SG SI
 SF SL TJ TM TR TT VA WG US VN YU ZA SW
 A# 2000043556 A 20001102 (200177) 9018831-00
 ADT WO 2000043687 A1 WO 2000-US10302 2000414; A# 2000043556 A AU 2000-43556
 20000414

FDT AU 2000043556 A Based on WO 200063687

PRAI US 1999-327383 19990608; U.S. 1999-129469 19990415

IC ICM GOING01-00

ICS G06F015-00; G06F017-00

AB WO 200063687 A UPAB: 20001209

NOVELTY - Identifying novel nucleic acids molecules encoding a **protein** of interest, using regulatory networks, is new.

DETAILED DESCRIPTION - Identifying novel nucleic acids molecules encoding a **protein** of interest, using regulatory networks, is new. The method comprises:

- (a) selecting a specific **protein** from a species involved in a regulatory network of interest;
- (b) identifying known **proteins** that act upstream and downstream of the **protein**, within the regulatory network;
- (c) constructing the regulatory network of interest from the **proteins** identified in (b);
- (d) for each identified **protein**, selecting a domain or motif and searching by homology for related **proteins** in a second species, a related **protein** has a homologous domain or motif;
- (e) producing a regulatory network for the second species, which incorporates the identified related **proteins**;
- (f) comparing the networks of the two species;
- (g) identifying a **protein** present in only one of the networks; and
- (h) isolating a nucleic acid molecule encoding the **protein** identified in (g) in the species in which it is missing.

INDEPENDENT CLAIMS are also included for the following:

- (1) identifying the effect of a gene knockout on a regulatory pathway, comprising:
 - (a) identifying the shortest non-oriented pathway connecting two gene products;
 - (b) assigning an initial sign value of minus to the knockout since the knockout gene is inactive;
 - (c) moving along the shortest pathway between the two gene products multiplying the sign with the sign of the next gene product in the pathway, where minus stands for inhibition and plus stands for induction or activation and zero stands for lack of interaction between two **proteins** in the specified direction; and
 - (d) determining the final sign at the end of the pathway, where minus indicates inhibition and plus indicates induction or activation of the pathway;
- (2) identifying a novel nucleic acid molecule encoding a **protein** of interest, comprising:
 - (a) selecting a gene of interest and searching a database for homologous sequences;
 - (b) aligning the sequences identified in (a);
 - (c) constructing a gene tree using the sequence alignment;
 - (d) constructing a species tree;
 - (e) inputting the species tree and gene tree into an algorithm which integrates the species tree and gene tree into a reconciled tree; and
 - (f) identifying orthologous genes present in one species but missing in another;
- (3) identifying a novel gene, comprising:
 - (a) defining a motif or domain composition of a gene of interest;
 - (b) searching for sequences which correspond to nucleotide sequences in an expression sequence tag database or other cDNA database using a program such as BLAST and retrieving the identified sequences;
 - (c) searching additional databases for expressed sequence tags containing the domains and motifs characteristic for the gene of interest with a hidden Markov model of domains and motifs identified in (A); and
 - (d) identifying nucleotide sequences comprising the gene of interest;
- (4) extracting information on interactions between biological entities from natural-language text data, comprising:
 - (a) parsing the text data to determine its grammatical structure; and
 - (b) regularizing the parsed text data to form structured word terms; and

(5) a computer system for extracting information on biological entities from natural-language text data, comprising:

- (a) means for parsing the natural-language text data; and
- (b) means for regularizing the parsed text data to form structured word terms.

USE - For identifying novel genes and for natural language processing and extraction of relational information associated with genes and proteins that are found in genomics journal articles.

ADVANTAGE - The method allows the rapid retrieval of information from literature and manipulation of derived functional data, removing a researcher's need to perform laborious reading and manual integration of research articles.

Iwg.0/23

FS CPI SPI
FA AB; PON
MC CPI: B04-E01; B11-C09; B12-K04E; D05-H09
EPI: S03-E03; T01-J

TECH UPTK: 20001209

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preferred Method: The nucleic acid encodes a human **protein**. The related **proteins** are orthologs. The regulatory pathway is involved in apoptosis. The specific **protein** from the first species is involved in tumor suppression.

TECHNOLOGY FOCUS - COMPUTING AND CONTROL - Preferred Method: In the method of (1), the following algorithm is used to integrate the species and gene tree into a reconciled tree:

- (a) computing the similarity $\sigma_{gi,ssj}$ for each pair of interior nodes from trees T_g and T_s ;
- (b) finding the maximum $\sigma_{gi,ssj}$;
- (c) saving sgi as a new cluster of orthologs, saving $(sgi)-(ssj)$ as a set of species that are likely to have a gene of this kind (or lost it in evolution);
- (d) eliminating sgi from T_g , $T_g := T_g / sgi$; and
- (e) repeating steps (b)-(d) until T_g is non-empty.

The method of (3) further comprises using each identified expression sequence tag to search sequence databases for overlapping sequences, to assemble longer overlapping stretches of DNA. The method of (4) further comprises preprocessing, the data prior to parsing. Preprocessing comprises identifying biological entities. The method further comprises referring to an additional parameter which indicates the degree to which subphrase parsing is to be carried out. The parsing step may comprise segmenting the text data by sentences and segmenting each sentence at identified words or phrases, or at a prefix. Alternatively, the parsing step comprises skipping undefined words or identifying binary actions and their relationships, and identifying arguments associated with the actions. The method may further comprise performing error recovery when parsing of the data is unsuccessful. The error recovery comprises segmenting the text data and analyzing the segmented data to achieve at least partial parsing of the unsuccessfully parsed data. The tagging step comprises providing the structured data component in a Standard Generalized Markup Language (SGML) compatible format.

Preferred System: The system further comprises means for preprocessing the data prior to parsing, and means for referring to an additional parameter which indicates the degree to which subphrase is to be carried out. The parsing means further comprise means for segmenting the text data into sentences and means for segmenting the sentences at identified words, phrases or prefixes. The parsing means may comprise means for skipping undefined words, means for identifying binary actions and their relationships, and means for identifying arguments associated with the actions. Means for performing error recovery when the parsing is unsuccessful may also be included. These comprise means for segmenting the text data and means for analyzing the segmented text data to achieve at least a partial parsing of the unsuccessfully parsed text data. Tagging means comprise means for providing the structured data component in SGML compatible format.

L123 ANSWER 5 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 0000-6409-9 [52] WPIX
 DNN NL 000-471469 DNC C.000-193184
 TI Estimating property values of sequences by the self-correlation of the property values, for predicting **protein** secondary structure .
 DC B04 D16 J04 T01
 FA (JEP-11) SEIBUTSU BUNSEI KOGAKU KENKYUSHO KK
 CYC 1
 P1 JP 1990-07378 A 20000728 (20062)* G06F015-18 <--
 ADT JP 199007378 A JP 1999-10964 19990119
 PRAI JP 1999-10964 19990119
 IC ICHI G06F015-18
 IC G06F017-18; G06F017-50
 ICA C06K14-00; C12N011-00
 AB JP200007378 A UPAB: 100001180
 NOVELTY - A method (M1) for estimating property values, by the self-correlation of the property values which determine the characteristics of the nature of a one-dimensional symbol string when a few of the particulars are placed in sequence, is new.
 DETAILED DESCRIPTION - In M1 when the one-dimensional symbol string is expressed as a sequence string depending on the property values, and two symbol strings and property values are given, then the property value is characterized by expressing the similarity of the sequence **profile** as the LPC cepstrum distance, and estimating it from the one-dimensional symbol string which can be constructed from symbols having unknown property values.
 An INDEPENDENT CLAIM is also included for estimating the self-correlation of the property values which determine the characteristics of the nature of a one-dimensional symbol string when a few of the particulars are placed in sequence. The most appropriate property value for the determination of the characteristics of its nature, is found from the native sequence, where a few of the particulars in sequence constitute the one-dimensional symbol string. This is done by finding the time sequence pattern between the one-dimensional symbol string, and symbols which artificially shuffled these randomly, and expressing as a sequence string of the shuffle sequence for the property values of each one-dimensional symbol string consisting of sequences having certain property values; then while comparing the similarity, by expressing it as the LPC cepstrum distance between two symbol strings, then estimating by a genetic algorithm so that the LPC cepstrum distance between the native sequence and the shuffle sequence becomes larger, and estimating its property value from the one-dimensional symbol string which can be constructed from symbols having unknown property values.
 USE - To determine the properties of amino acid sequences of **proteins** and the base sequences of genes.
 ADVANTAGE - The amino-acid index can be estimated from the sequence data only. It is similar to the hydrophobic index, giving the influence of **protein** folding, and the differences in the three-dimensional structure of the **protein** when there is a good deal of secondary **protein** structure.
 Dwg.0/0
 FS CPI EPI
 FA AB; DCN
 MC CPI: B04-N04; B10-B02; B11-C08B; B12-K04; B05-B09; J04-C03
 EPI: T01-J03; T01-J15; T01-J16
 L123 ANSWER 6 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 0000-594572 [56] WPIX
 DNN NL 000-440291 DNC C.000-177635
 TI Monitoring differential expression of genes in filamentous fungal cells uses fluorescence-labeled nucleic acids isolated from the cells and a substrate of expressed sequence tags.
 DC B04 C06 D16 T01
 IN BERKA, R M; CLAUSEN, I G; KAUPPINEN, S; OLSEN, P B; REY, M W; SHUSTER, J R
 PA (NOVO) NOVO NORDISK AS; (NOVO) NOVO NORDISK BIOTECH INC
 CYC 6*

PT WO 2000056762 A2 20000328 (2000056)* EN C07K014-37 --
 RW: AT BE CH CY DE IF EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL
 CA PT SD SE SI SJ TZ UG ZW
 W: AE AL AM AT AU AZ BA BB BG BP BY CA CH CN CR CU CZ DE DK DM BE ES
 FI GB GE GH GM HP HU ID IL IM IS JP PE KG KP KP KZ LC LF LS
 LT LU LV MA ME MI MF MN MW MX NP NS PI PT EG RI SP SE SG SI SF SL
 TJ TM TR TT UA UG UK VH YU ZA ZW

AU 2000039154 A 2000-100-9 (200103) C07K014-37 --
 ADT WO 2000056762 A2 WO 2000-037781 2000-0322; AU 2000039154 A AU 2000-39154
 20000322

FDT AU 2000039154 A Based on WO 200056762

PRAI US 1919-173622 15-01-2002

IC C07K014-37

EGS C06F001-68; G06F017-30; G06F019-00

AB WO 2000056762 A UPAT: 20011106
 NOVELTY - New method for monitoring differential expression of genes in a first filamentous fungal (FF) cell relative to expression of the same genes in one or more second filamentous fungal cells uses fluorescence-labeled nucleic acids isolated from the FF cells and a substrate of expressed sequence tags (EST).

DETAILED DESCRIPTION - Method for monitoring differential expression of genes in a first filamentous fungal (FF) cell relative to expression of the same genes in one or more second FF cells comprising:

(1) adding a mixture of fluorescence-labeled nucleic acids isolated from the FF cells to a substrate containing an array of FF expressed sequence tags (EST) in the array, where the nucleic acids from the first and one or more second FF cells are labeled with different first and second fluorescent reporters respectively; and

(2) examining the array by fluorescence under fluorescence excitation conditions where the relative expression of the genes in the FF cells is determined by the observed fluorescence emission color of each spot in the array in which:

(i) the ESTs in the array that hybridize to the nucleic acids obtained from either the first or the one or more second FF cells produce a distinct first fluorescence emission color or one or more second fluorescence emission colors, respectively; and

(ii) the ESTs in the array that hybridize to the nucleic acids obtained from both the first and the one or more second FF cells produce a distinct combined fluorescence emission color.

INDEPENDENT CLAIMS are also included for the following:

(1) a computer readable medium which has recorded on it an array of FF ESTs for monitoring differential expression of genes in a first FF cell relative to expression of the same genes in one or more second FF cells;

(2) a computer-based system for monitoring differential expression of genes in a first FF cell relative to expression of the same genes in one or more second FF cells comprising:

(a) a data storage device comprising FF ESTs selected from sequences 1-7860 (given in the specification), nucleic acid fragments of sequences 1-7860 or nucleic acid sequences with at least 50% homology to sequences 1-7860;

(b) a search device for comparing a target sequence to a FF EST sequence of the data storage device of (a) to identify homologous sequences; and

(c) a retrieval device for obtaining the homologous sequence of (b);

(3) a substrate comprising an array of FF ESTs for monitoring differential expression of genes in a first FF cell relative to expression of the same genes in one or more second FF cells; and

(4) an isolated nucleic acid sequence comprising any of sequences 1-7860.

USE - The ESTs are used in the methods for monitoring differential expression of genes in a first filamentous fungal (FF) cell relative to expression of the same genes in one or more second filamentous fungal cells (claimed). Monitoring the global expression of genes from FF cells allows the production potential of the microorganisms to be improved. New genes may be discovered, possible **functions** of unknown open reading frames can be identified and gene copy number variation and

stability can be monitored. The expression of genes can be used to study how FF cells adapt to changes in culture conditions, environmental stress, spore morphogenesis, recombination, metabolic or catabolic pathway engineering.

ADVANTAGE - Using ESTs provides several advantages over **genomic** or random cDNA clones including elimination of redundancy as one spot on an array equals one gene or open reading frame, and organization of the microarrays based on **function** of the gene products to facilitate analysis of the results.

Twg.0/C

FS CPI EPI

FA AF; DCM

MC CPI: B04-E03; B04-E05; B04-F09A; B11-C07B3; B11-C08E5; **B12-K04E**; **B12-K04F**; C04-E03; C04-F05; C11-F09A; C11-C07B3; C11-C08E5;

C12-K04E; **C12-K04F**; D04-H05; D05-H09;

D05-H10; D05-H12A; D05-H12B

EPI: T01-J; T01-J05B

TECHN UPTK: 20001106

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preferred Sequences: The ESTs are *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolypocladium* or *Trichoderma* ESTs. The *Fusarium venenatum* ESTs are selected from sequences 1-3770 (given in the specification), nucleic acid fragments of sequences 1-3770 or nucleic acid sequences with at least 90%, preferably 95% and especially 99.9% homology to sequences 1-3770. The *Aspergillus niger* ESTs are selected from sequences 3771-4376 (given in the specification), nucleic acid fragments of sequences 3771-4376 or nucleic acid sequences with at least 90%, preferably 95% and especially 99.9% homology to sequences 3771-4376. The *Aspergillus oryzae* ESTs are selected from sequences 4377-7401 (given in the specification), nucleic acid fragments of sequences 4377-7401 or nucleic acid sequences with at least 90%, preferably 95% and especially 99.9% homology to sequences 4377-7401. The *Trichoderma reesei* ESTs are selected from sequences 7402-7860 (given in the specification), nucleic acid fragments of sequences 7402-7860 or nucleic acid sequences with at least 90%, preferably 95% and especially 99.9% homology to sequences 7402-7860.

Preferred Method: The hybridization conditions are very low, low, low-medium, medium, medium-high, high or very high stringency conditions.

Preferred Substrate: The substrate comprising an array of FF ESTs may be a glass support with hydrophobic or hydrophilic coating on its surface to which the ESTs are electrostatically bound non-covalently, with each EST at a separate, defined position. The substrate may be a multicell substrate with each cell containing a microarray of ESTs.

TECHNOLOGY FOCUS - BIOLOGY - Preferred Cells: The FF cells are *Acremonium*, *Aspergillus*, *Fusarium*, *Humicella*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolypocladium* or *Trichoderma* cells. The 2 or more FF cells are the same cell or different cell, preferably *Fusarium venenatum*, *Aspergillus niger* or *Aspergillus oryzae* cells.

TECHNOLOGY FOCUS - COMPUTING AND CONTROL - Preferred Medium: The computer readable medium is a floppy disk, a hard disk, random access memory (RAM), read only memory (ROM) or CDROM.

L123 ANSWER 7 0F 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 2000-573798 [54] WPIX

DNN N2000-424559 ENC C2000-171175

TI Clustering gene expression datapoints in a computer system using a self-organizing map.

DC B14 S16 T01

IN GULBZ, T F; LANGER, E S; MESIPOV, J; TAMAYO, P

PA (WHED) WHITEHEAD INST BIOMEDICAL FES

CYC 27

PI EP 1037158 A2 20000920 (200054)* EN 39p G06F019-00

P: AL AT BE CH CY DE DK ES FI FR GB GF IE IT LI LT LU LV MC MK NL PT

PO SE SI

CA 2300639 A1 20000915 (200056) EN G06F019-00 <--
 JP 200034239 A 20001212 (200104) 34p C12Q001-68

ADT EP 1037158 A2 EP 2000-302029 20000314; CA 2300639 A1 CA 2000-2300639
 20000314; JP 200034239 A JP 2000-73166 20000315

PRAI US 1999-14453 13040315

IC ICM C12Q001-68; G06F019-00

ICS C12M001-00; C12N001-00; C12N015-09; G01N033-15; G01N033-50

ICA G06F017-39

AB EP 1037158 A (PCT: 19991027)

NOVELTY - A method for clustering gene expression datapoints in a computer system using a self-organizing map, is new.

DETAILED DESCRIPTION - Method for clustering datapoints (each datapoint is a series of gene expression values) in a computer system, comprising:

- (i) receiving the gene expression values of the datapoints;
- (ii) using a self-organizing map (SOM), clustering the datapoints so that datapoints that exhibit similar patterns are clustered together into respective clusters; and
- (iii) providing an output indicating the clusters of the datapoints.

INDEPENDENT CLAIMS are also included for the following:

- (1) a method for grouping datapoints in a computer system, where each datapoint is a series of gene expression values, comprising:
- (i) receiving gene expression values of the datapoints;
- (ii) filtering out any datapoint's that exhibit an insignificant change in the gene expression value, so that working datapoints remain;
- (iii) normalizing the gene expression value of the working datapoints;
- (iv) using a SOM, grouping the working datapoints so that datapoints that exhibit similar patterns are grouped together into respective clusters; and
- (v) providing an output indicating the groups of the datapoints;
- (2) a computer apparatus for clustering datapoints, where each datapoint is a series of gene expression values, comprising:
- (i) a source of gene expression values of the datapoints;
- (ii) a processor routine coupled to receive datapoints from the source, the processor routine utilizing a SOM for clustering datapoints so that datapoints that exhibit similar patterns are clustered together into respective clusters; and
- (iii) an output device, coupled to the processor routine, for indicating the clusters of datapoints;
- (3) a computer apparatus for grouping datapoints, where each datapoint is a series of gene expression values, comprising:
- (i) a source of gene expression values of the datapoints;
- (ii) a filter coupled to the source, for receiving the gene expression values and filtering out any of the datapoints that exhibit an insignificant change in the gene expression value, so that working datapoints remain;
- (iii) a normalizing process, coupled to the filter, for normalizing the gene expression value of the working datapoints;
- (iv) a processor routine that is responsive to the normalizing process and utilizes a SOM for grouping the working datapoints such that datapoints that exhibit similar patterns are grouped together into respective groups; and
- (v) an output device, coupled to the processor routine, for indicating the clusters of datapoints;
- (4) a method for assessing expression patterns of two or more genes in cells, where the expression patterns are represented by datapoints, and each datapoint is a series of gene expression values, comprising:
- (i) receiving the gene expression values of the datapoints;
- (ii) using a SOM, clustering the datapoints such that datapoints that exhibit similar patterns are clustered together into respective clusters;
- (iii) providing an output indicating the clusters of datapoints; and
- (iv) analyzing the output to determine the similarities or differences between the expression patterns of the genes;
- (5) a method of determining relatedness of expression patterns of two

or more genes, where the expression patterns are represented by datapoints and each datapoint is a series of gene expression values, comprising:

- (i) receiving the gene expression values of the datapoints;
- (ii) using a SOM, clustering the datapoints such that datapoints that exhibit similar patterns are clustered together into respective clusters;
- (iii) providing an output indicating the clusters of datapoints; and
- (iv) analyzing the output to determine the similarities and/or differences between the expression patterns of the genes, thereby determining the relatedness of the genes;

(6) a method for characterizing expression patterns of genes of a sample having unknown characteristics, where the sample is obtained from an individual and subjected to diagnostic tests, and the expression patterns of the genes for the diagnostic tests are represented by datapoints, and each datapoint is a series of gene expression values across multiple genes for the diagnostic test, comprising:

- (i) receiving the gene expression values of the datapoints from the diagnostic tests;
- (ii) using a SOM, clustering the datapoints such that datapoints that exhibit similar patterns are clustered together into respective clusters;
- (iii) providing an output indicating the clusters of datapoints; and
- (iv) comparing the output of the gene expression patterns of the unknown sample against a control, thereby characterizing gene expression patterns of the sample;

(7) a method of identifying a drug target from the expression patterns of two or more genes from cells, where the expression patterns are represented by datapoints and each datapoint is a series of gene expression values, comprising:

- (i) obtaining cells that express genes;
- (ii) subjecting the cells to an agent or condition for testing the drug target;
- (iii) measuring gene expression from the cells subjected to the agent or condition, and from a control, to obtain the gene expression values;
- (iv) receiving the gene expression values of the datapoints;
- (v) using a SOM, clustering the datapoints such that datapoints that exhibit similar patterns are clustered together into respective clusters;
- (vi) comparing the clusters from the genes that have been subjected to the agents or condition with a control; and
- (vii) providing an output indicating clusters, to thereby determine the drug target;

(8) a drug target identified or identifiable by the method of (7);

(9) a computer-readable product on which is recorded a program loadable into the internal memory of a digital computer and comprising software code portions for performing the steps of the above methods.

USE - The method can be used, e.g. to identify drug targets from the expression patterns of two or more genes and to analyze the relatedness of two or more genes, the unknown function of a gene under known conditions, the effect of unknown conditions on a known gene function or the likelihood of successful treatment by a drug (e.g. for a specific tissue sample).

ADVANTAGE - Using SOMs to cluster gene expression patterns into groups exhibiting similar patterns makes it easy to analyze gene expression data from potentially thousands of genes.

DESCRIPTION OF DRAWING(S) - The figure is a schematic diagram illustrating the principle behind the self-organizing map, in which the initial geometry of nodes in a 3x2 rectangular grid is indicated by solid lines connecting the nodes, datapoints are represented by black dots, the nodes are represented by large circles, and trajectories are represented by arrows.

Dwg.1/6

FS	CPI	EPI	
FA	AB;	SI;	DCN
MC	CPI: B04-C01; B04-E01; B04-N04; B11-C08E1; B11-C09;		
	B12-K04A; B12-K04E; D05-H02; D05-H09;		
	D05-H11; D05-H17; D05-H18; D05-J		
	EPI: T01-J		

L123 ANSWER 8 OF 73 WPIX COPYRIGHT 2001 DEPARTMENT INFORMATION LTD

AN 2000-27782 [54] WPIX

DNN 02500-424943 DNC 02000-171170

TI Detection of WT1 gene **functional** mutations and EWS-WT1 gene fusion's by monitoring downstream genes regulated by WT1, useful for diagnosis of neoplasia and for anti-cancer drug screening.

DC R04 D16 S05 T01

IN HABEE, D; LEE, S; CLINEP, J; TPHONG, V

PA AFFY-N, AFFYMETRIX INC

CYC 26

PI EP 1036645 A2 20000920 (20064) EN 23p C12N001-68
E: AI AT BE CH CY DE DK ES FI FR GR IE IT LI LT LU LV MC MK NL PT
RO SE SI

CA 2296782 A1 20000824 (20064) EN C12N001-68

EP 000164560 A 20011007 (20064) ***P C12N001-68

US 6177248 P1 20010122 (20017) C12N001-68

ADT EP 1036645 A2 EP 2000-10-262 20000215; CA 2296782 A1 CA 2000-2296782
20000210; JP 2000368600 A JP 2001-64155 20000202; US 6177248 B1 US
1990-156301 19900224

PRA1 US 1990-156301 19900224

IC ICM C12N001-68

PCT C07H021-01; C12N005-10; C12N015-00; C12N015-00; G01N033-15;
G01N033-00; G01N033-05; G01N033-066; G01N033-074; G06F019-00

ICI C12N005-10; C12N001-91

AB EP 1036645 A UEPAT: 2001027

NOVELTY - A method for the detection of WT1 gene **functional** mutations or an EWS-WT1 gene fusion by monitoring expression levels of downstream genes, which are regulated by wild type WT1 and/or EWS genes, is new.DETAILS DESCRIPTION - Methods for detecting a WT1 gene **functional** mutations or EWS-WT1 gene fusions in target cells comprises:

(1) detecting expression of at least 3 downstream genes of WT1 or one or more downstream genes of an EWS-WT1 fusion in a sample of target cells and reference cells having a wild type WT1 and/or EWS gene (the reference cells are otherwise substantially similar to the target cells, the downstream genes are up- or down-regulated by the wild type WT1 and/or EWS gene); and

(2) comparing the expression of the downstream genes in the target cells and the reference cells (a difference in the expression between the target cells and reference cells suggests a WT1 **functional** mutation or EWS-WT1 fusion in the target cells).

INDEPENDENT CLAIMS are also included for the following:

(a) an in-cell **functional** assay for a WT1 sequence alteration comprising a method as above;

(b) methods for detecting a mutation in a target WT1 gene or a translocation fusing EWS and WT1 genes using a computer;

(c) a method of diagnosing neoplasia of a test cell comprising:

(i) hybridizing a transcription indicator of a test cell to a set of nucleic acid probes (the transcription indicator is chosen from mRNA, cDNA and cRNA, the probes comprise a number of nucleic acid molecules each of which is a portion of a gene which is activated by or repressed by WT1 or EWS-WT1 fusion **protein**);

(ii) detecting amounts of transcription indicator which hybridize to each of the set of probes;

(iii) identifying a test cell as neoplastic if hybridization of the transcription indicator of the test cell to a probe which is a WT1- or EWS-WT1-activated gene is lower than hybridization using a transcription indicator from a normal cell, or hybridization of the transcription indicator of the test cell to a probe which a WT1- or EWS-WT1-repressed gene is higher than hybridization using a transcription indicator from a normal cell;

(d) a method of identifying anti-cancer drugs comprising:

(i) contacting a test compound with a human cell;

(ii) determining the expression of the cell of at least one WT1 or EWS-WT1 up- or down-regulated gene; and

(iii) identifying a test compound as a potential anti-cancer drug if it increases expression of at least one WT1 or EWS-WT1 up-regulated gene or decreases expression of a WT1 or EWS-WT1 down-regulated gene in the human cell.

USE - The methods can be used to diagnose cancers, for drug-screening and to **functionally** analyze mutations involving the WT1 gene. The methods use probes based on genes identified as being regulated by WT1 and/or WT1 fusions to EWS. The methods can be used to determine the genetic status of the gene and to identify which have similar effects on downstream genes.

Dwg.070

FS
FA
MC

EPI: B04-C01; B04-E01; B04-E01F; B04-F01; B04-F9100E; B04-H01;
B04-N04; B11-C06H1; B11-T04; **B12-K04A3**;

B12-K04E; D05-H04; **D05-H09**; D05-H12A; D05-H12B;
D05-H14; D05-H17A; D05-H17C; D05-H17

EPI: S05-C; T01-J

TECH

PTX: 20001027

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preferred Methods: The downstream genes are transcriptionally regulated by the wild type WT1 gene and the expression of the downstream genes is detected by measuring amounts of transcripts of the downstream genes in the reference and target cells. Amounts of transcripts are measured with a high-density nucleic acid array. The downstream genes comprise at least one WT1 up-regulated gene chosen from natural killer cells **protein 4 precursor** (M26507), folate binding **protein** (M25317), HALPHA44 gene for alpha-tubulin (X06956), heat shock **protein** HSP70B (X51751), 90K product (H17960), heat shock 70 kD **protein** 1 (T66317), amphiregulin (M30704), procollagen-alpha 1 (P51851), beta-migrating plasminogen activator inhibitor 1 (M14863), cysteine protease inhibitor from radiated keratinocytes (X05375), brain natriuretic **protein** (M1766), leukemia virus receptor 1 (GLVR1) (M20501), type 1 cytoskeletal 17 keratin (F71870), tubulin alpha-5 chain (H45051), purine nucleoside phosphorylase (T47964), GMP-GTPase (U10550), adrenomedullin (D14674), GPI-anchored urokinase plasminogen activator surface receptor (F36636), tissue type plasminogen activator (X13097), gravin (M96122), jun-B (X51343), elongation factor 1 alpha-2 (X79430), homeotic gene regulator (R16977), clone 9112 (X57348), interferon gamma treatment inducible mRNA (M26683), transmembrane receptor **protein** (31727), acidic FGF (X65779), cartilage linking **protein** 1 (X17405), interleukin (IL-11) (X58377) and mitochondrial phosphate carrier **protein** (R49231).

The downstream genes comprise at least one WT1 down-regulated gene chosen from caveolin and elongation factor 1 alpha-2. The downstream genes further comprise p11 or EGFR (endothelial growth factor).

The method further comprises indicating the loss of **function** mutation in the WT1 gene in the target cells if the expression of WT1 up-/down-regulated genes is at least two times less/more in the target cells than in the reference cells.

The downstream genes comprise at least one EWS-WT1 up-regulated gene chosen from Pseudorabies virus nuclear antigen (R60406 and T79475), beta chain of IL-2 receptor (M26062), G1/S-specific cyclin D1 (H20529), neuron-specific gamma-2 enolase (M22840), fusion gene NPM-MLF1 resulting from translocation t(3;5)(q25.1;p34) (L4.0084), collagen alpha1(IV) chain (H62466), T-lymphocyte specific **protein** tyrosine kinase p65ck(18k) (U23911), insulin-like growth factor binding **protein** 4 (U10981), cathepsin B (L16510), cystatin C (T51534), dual specific mitogen-activated **protein** kinase J (F42241), neuronal-glial cell adhesion molecule (T53118), FIM-1 proto-oncogene serine/threonine **protein** kinase (H14925), carboxylesterase precursor (F54359), **protein**-tyrosine phosphatase PAC-1 (L11129) and interferon-induced **protein** 6-16 (X02432).

Necplasia in kidney cells or Ewing sarcoma cells can be detected using probes of genes, which are activated or repressed by WT1 or EWS-WT1, respectively. The probes comprise portions of the genes and are preferably attached to a solid support and arranged in an array. Preferably, at least

6000 different genes are used as probes. The WTI genes in a test cell are sequenced to determine the WTI genotypic status of the cell.

L123 ANSWER 3 OF 73 WPIX COPYRIGHT 2001 PERWENT INFORMATION LTD
 AN 2000-572101 [53] WPIX
 DNN N2000-413219 DNC C2000-170570
 TI Identifying **peptide** with selected **function**, useful particularly for C-amidated hormones, by screening database for combination of nucleic acid and amino acid sequences.
 DC B04 E16 T01
 IN BERGE, G; CAMAFA Y FEFER, C A; GOZE, C; MARTINEZ, J; THUREAU, C
 FA CNRS CENT NAT RECH SCI; EURP PCMAS SOC CONCILIS RECH & APPL SCI
 CYC 47
 PI WO 2000050636 A1 2000041 (00003) FR 19p C13Q001-68
 RW: AT BE CH CY DE DK ES FI FR GP GR GM GR IE IT KE LS LU MC MW NL
 OA PT SD SE SI SG TG UG SW
 W: AE AL AM AT AU AS BA BB BG BR BY CA CH CN CU CG DE DK EE ES FI GB
 GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KU LC LK LE LS LT LU
 LY MD MG MK MN MW ME NO NZ PL PT RO RU SD SE SG SI SK SI TJ TM TR
 TT UA UG US UZ VI YU SA SW
 AU 2000029209 A 20000314 (200003) C13Q001-68
 ADT WO 2000050636 A1 WO 2000-FR460 P 000024; AU 20000314 A AU 2000-29209
 30000224
 FDT AU 2000029209 A Based on WO 2000050636
 PRAI US 1999-207625 19990128
 IC ICM C13Q001-68
 P3 G06F017-30
 AB WO 2000050636 A UPAE: 20001023
 NOVELTY - Identifying a **peptide** (I) having a particular **function** comprising preparing a database of polynucleotid-s (PN) and **polypeptides** (PP) of unknown **functions**, screening the database for a combination (A) of nucleotides or amino acids indicative of (I) with a particular **function**, and identifying PN and PP which contain (A), is new.
 DETAILED DESCRIPTION - An INDEPENDENT CLAIM is also included for a similar method for identifying **peptide** precursors (II) having an amidated C-terminus.
 USE - The method is used to identify precursor **peptides** with an amidated C-terminus, especially **polypeptide** hormones, for studying physiologically active substances.
 ADVANTAGE - The method is not limited to (I) expressed by a particular strain, does not depend on physical measurements (isoelectric point or solubility), can be applied to **proteins** which are not apparently related by physical properties, and can identify candidate (I) with very low mutual homology. It can be performed with readily available databases and computers.
 Dwg.0/0
 FS CPI EPI
 FA AB; DCN
 MC CPI: B04-E01; B04-N04; B11-C08; B12-K04E;
 B12-K04F; D05-H09
 EPI: T01-105B
 TECH UPTK: 20001023
 TECHNOLOGY FOCUS - BIOLOGY - Preferred Process: To identify (II), the database contains both PN and corresponding PP sequences, especially with PP being derived by translation from PN, and each PN is translated into three PP (one for each reading frame). The database also includes annotations related to the sequence, especially origin, source, characteristics and references. The screening step additionally includes location of the AUG start codon and verification that no stop codon is present between this AUG and the end of the coding region for (II). Once PP/PN have been identified, they are compared with sequences of known **function** and those with unknown **function** are retained. These are then examined for similarity with sequences of known **function**, and those for which no similarity can be found are selected for further analysis, or as candidate sequences. Optionally the

selected FP sequences are synthesized and their properties evaluated.

Preferred Materials: To identify (II), the combination of nucleotides is

Y1-Y2-Y3-Y4-Y5

Y1 = 0-11 nucleotides;

Y2 = codon for Gly;

Y3 and Y4 = codons for Arg or Lys;

Y5 = 0-11 nucleotides.

L123 ANSWER 10.0F /3 WFIG COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 2000-05-426 (S1) WFIG
 DNN B2000-413204 IHC C2000-160375
 TI Identifying mutations, useful e.g. for diagnosing cancer or predisposition to it, by mathematical analysis of output patterns from DNA microarrays.
 DC P04 D16 M04 T01
 IN CHIATI, S
 PA VITAL-II VITALOGY CORP
 CYC 96
 PI WO 2000051056 A2 20000831 (2000-1) EN 37p G06F019-00 --
 FW: AT BE CH CY DE DF EA EG FI FF GB GH GM GR IE IT KE LS LU MC MW NL
 OA PT SD SE SL SS TS UG SW
 W: AE AL AM AT AU AS BA BB BG BF BY CA CH CN CP CU CZ DE DK DM EE ES
 FI GB GD GE GH GM HF HU ID IL IN IS JP KE KG KP KF KZ LC IK LR LS
 LT LU LV MA ME MG MN MW MX NO NZ PL PT RU SU SE SG SI SK SL
 TI TM TR TT TS VA UG UC VN YU SA SW
 US 614,681 A 20000107 (2000-1) G06F019-18 --
 AU 200003686 A 20000914 (2000-3) G06F019-00 --
 ADT WO 2000051056 A2 WO 2000-051056-0217; US 61436-1 A US 1999-293792
 19990727; AU 200003686 A AU 2000-03686 2000-0217
 FDT AU 200003686 A Based on WO 200051056
 PPAI US 1999-293792 19990221
 IC ICM G06F015-18; G06F019-00
 AB WO 2000051056 A UPAE: 200001016
 NOVELTY - A biochip output pattern (OP) is analyzed to identify mutations in biological samples applied to it. The OP is tessellated and a stimulus pattern (SP) associated with known mutations is generated. A resonance pattern (FP), of resonance between SP and tessellated OP, is generated and interpreted to yield confirmed mutations.
 DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:
 (1) preconditioning the OP of a biochip, comprising:
 (a) tessellating the output pattern to match a predetermined stimulus pattern yielding a tessellated output pattern;
 (b) extracting local parametrics from the tessellated output pattern;
 (c) determining if a degree of amplitude wandering representative of the local parametrics is within an allowable generator function limit; and
 (d) if not, renormalizing the tessellated output pattern to match spectral properties of the stimulus pattern and repeating (b)-(d);
 (2) performing a convergent reverberant dynamics resonance analysis on the OP of a biochip to identify mutations, comprising:
 (a) determining resonance dynamics relaxation values based on the preconditioned output pattern and resonance stimulus;
 (b) filtering the dynamics relaxation values using ensemble boundary and CSR filters to yield a second set of values;
 (c) applying bulk property estimators to the dynamics relaxation values to yield a third set of values;
 (d) evaluating the second and third sets of values to determine if resonance convergence has been achieved; and
 (e) determining if dynamic paralysis has occurred, and repeating the method if it has;
 (3) system for analysing OP of a biochip to identify mutations comprising:
 (a) a tessellation unit operative to tessellate the output;
 (b) a stimulus pattern generation unit;
 (c) a resonance pattern generation unit and
 (d) a resonance pattern interpretation unit; and

(4) system for performing method (3), comprising:
 (a) a determination unit;
 (b) a filter to filter dynamics relaxation values using ensemble boundary and CSP filters;
 (c) an estimator unit;
 (d) an evaluation unit; and
 (e) a paralysis detection unit.

USE - The method is used to identify mutations, particularly for diagnosis of disease (e.g. cancer and acquired immune deficiency syndrome) and to identify susceptibility to disease.

ADVANTAGE - Tessellation of OP improves the repeatability of the analysis.

Swg.0/4

FS CPI EPI
 FA AP; DMN
 MC CPI: B04-E01; B04-N04; B11-C03; B12-K04A1;
 D05-H09; D05-H14; D04-C03
 EPI: T01-J

TECH UPTK: 200001016

TECHNOLOGY FOCUS - INSTRUMENTATION AND TESTING - Preferred Process: OP is determined from a microarray comprising many cells each with a set of identical immobilized oligonucleotides, different from those in all other cells, particularly a dot spectrogram. Tessellation comprises amplifying highly local morphological variations in OP, and generation of SP is based on Quantum Expressor **Functions** (QEF). OP is tessellated to match the morphological characteristics of QEF. Optionally local parametrics (LP) are isolated from the tessellated OP and a degree of amplitude wandering, representative of LP, is defined. If this degree is not within a predetermined allowable generator **function** limit, the tessellated OP is renormalized to match further spectral properties of the resonance pattern. If an representative of an integrated density of states within each region of the tessellated OP, and the degree of amplitude wandering is determined by applying a Falm distribution to develop generators for estimation of stochastic wandering. Optionally the tessellated OP and SP are transformed to a metrically transitive random field, optionally followed by renormalization of the tessellated OP. Generation of SP involves selection of a subset of mutations and selection of a subset of nonlinear QEF, then transforming the QEF and tessellated OP into phase space. PP is generated by iterative processing of OP by performing a convergent reverberation to produce a resonance pattern that represents resonance between a selected QEF and the tessellated OP until a predetermined degree of convergence (DC) is achieved between resonance in PP and that expected for the set of mutations. Optionally the resonance dynamics analysis method includes filtering the diagnostic conditions identified by the convergent reverberant dynamics analysis based on clustering properties, and a diagnostic decision is made based on these conditions. If the decision is negative, the availability of an alternative is determined, and if none is available then a new subset of mutations is chosen. The biological sample is DNA, RNA, **proteins**, **peptide**-nucleic acid, or samples from targeted nucleic acid amplification.

L123 ANSWER 11 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 2000-558335 [51] WPIX

CNN NJ 000-413159 INC CR000-16C301

II Dynamic storage, retrieval and analysis of experimental data e.g. cell biological studies and drug discovery using a graphical user interface to represent biological entities and relationships.

DC B04 D16 T01

IN HARRINGTON, C C; QU, L; TAYLOR, D L; WANG, J

PA CELL-N) CELLCOMICS INC

CYC 90

FI WO 00049540 A1 20000614 (200001)* EN 79p G06F017-50

FW: AT BE CH CY LE DK EA ES FI FF SP GH GM GR IE IT KE LS IC MC MW NL
 OA PT SD SE SL SE TS UG SW

W: AE AL AM AT AU AE BA BB BG BR BY CA CH CN CF CU CZ IE DK DM EE ES

FI GR GD GE GH GM HF HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS
 LT LU LV MA MD MG MF MN MW MX ND NZ PL PT RO RU SD SE SG SI SK SL
 TJ TM TR TT TZ WA UC US UC VN YU ZA ZW

AU 2000030027 A 20000994 (200103) G06F017-50
 ADT WO 2000049540 A1 WO 2000-03437/1 20000218; AU 2000030027 A AB 2000-30027
 20000218

FDT AU 2000030027 A Based on WO 200049540

PRAI US 1399-100-01 19990219

IC ICM G06F017-50

ICS G06F017-30; G06F019-00

ICI G06F15/00

AB WO 20004954 A UPAB: 20000218

NOVELTY - A method for dynamic storage, retrieval and display of experimental information with determined relationships, is new.

The method uses a graphical user interface (64) from which shapes (30, 34, 96, 98, 100, 101) and arrows (92, 104, 106, 108, 110, 112) representing biological entities and transformations respectively, can be input and edited. Multi-dimensional information based on a predetermined, but expandable, hierarchy is input to relate the information to additional information, e.g. from the Internet.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(i) a method for displaying experimental information with determined relationships from a remote computer; and

(ii) a system for dynamically storing retrieving and displaying of experimental information with determined relationships.

USE - The method and system is useful for storing, retrieving and analyzing experimental information for use in cell biological studies and drug discovery.

ADVANTAGE - The system allows easy storage, retrieval and analysis of biological information associated with biological pathways. It can access hierarchical information associated with biological pathways from multiple private and public databases.

DESCRIPTION OF DRAWING(S) - The figure is a block diagram showing a screen display of a graphical user interface used to create, store and analyse biological pathways with determined relationships.

graphical button for selecting a shape 66

graphical button for selecting an arrow 68

graphical button for selecting a cell organelle or compartment 70

outline of a cell 72

an outline of a cell membrane 76

other graphical buttons 75, 80, 82, 84, 86, 88

Dwg. 3/11

FS CPI EPI

FA AB; GI

MC CPI: B04-F01; B11-C08; B12-K04; D05-H09

EPI: T01-J05B2B; T01-J05B4F; T01-J05B4P; T01-J06A1; T01-J15X

TECH UPTK: 20001016

TECHNOLOGY FOCUS - INSTRUMENTATION AND TESTING - Preferred Method:

Information associated with the series of shapes interconnected with arrows as a biological pathway with determined relationships, is saved in a database. The pathway defines a hierarchical representation of a biological function with determined relationships between the entities and transformations. A computer readable medium stores instructions causing a central processing unit to execute the method. The biological entity represented by a shape may be a cell, a sub-component of a cell or an aggregation of a number of cells. The entities may include active or inactive entities, entity inhibitors, factors exchanged between entities or intermediate entity transformation products. The arrows represent biological transformations between entities. The transformations may include transcription factor activation, cellular hypertrophy, protein kinase activation, protease activation, gene expression, receptor activation, apoptosis, internalization of cell surface receptor proteins, mitochondrial potential, neurite outgrowth, cell viability or mitotic index for a sub-component of a cell, a cell or cell aggregation. Additional information inputted may be a component view, a

morphology of the entity, an electron microscope photograph, a fluorescent view of the entity, basic information, site information, **function** information, enzyme information, reaction information, transport system information or a pathway view.

L123 ANSWER TO QF 73 WPIX COPYRIGHT 2001 DEPWENT INFORMATION LTD
 AN 2000-548066 (50) WPIX
 CR 2000-570287 (50)
 DNN NL000-436520 DNC C2000-164107
 TI Unencrypting traits using splice gene sequences.
 DC P04 C06 P16 T01
 IN BERMUDES, E R; CAPP, E; LAMMER, M; NESS, J F; PATTEN, P A; YAMAMOTO, T
 PA MAXY-MI MACTGEN INC
 CYC 00
 PI WO 2000052146 AU 10000003 (2000) * EN 76p C12N009-00
 PW: AT RE CH CY DE DE RA ES FI FF GB GH GM GR IE IT HE LS LU MC MW NL
 OA FT SD SE SL SC TD UG CW
 W: AE AL AM AT AU AZ BA BB BG BF BY CA CH CN CP CU CZ DE DK DM EE ES
 FI GB GD GE GH GM HF HU TD TL IN IS JP KE KG KP KR KZ LC LK LR LS
 LT LU LV MA MD MG ME MN MW NO NZ PL PT RU SD SE SG SI SK SL
 TI TM TR TT TS UA UG US US VI YU ZA SW
 AU 2000033919 A (2000) 005 C12N009-00
 ADT WO 2000052146 A2 WO 2000-033919 (2000) 003; AU 2000033919 A AU 2000-33919
 20000303
 FDT AU 2000033919 A Based on WO 200052146
 PRAI US 1999-164618 19991110; US 1999-122943 19990305; US 1999-142299
 19990701; US 1999-164617 19991110
 IC I0M C12N009-00
 ICS A01H001-00; A01H005-00; C12N009-14; C12N015-10; C12N015-62;
 C12N015-62; G06F017-50
 AB WO 2000052146 A (PAPR: 2000) 14
 NOVELTY - Methods of unencrypting trait encrypted gene sequences to
 provide unencrypted RNA and **proteins**, are new.
 DETAILED DESCRIPTION - INDEPENDENT CLAIMS are included for the
 following:
 (1) a method (I) of unencrypting trait encrypted gene sequences to
 provide at least 1 unencrypted RNA or **polypeptide**, comprising:
 (a) providing a number of split gene sequences (Grp1) each of which
 comprises a subsequence of a genetic element;
 (b) transcribing the split gene sequences to provide a number of RNA
 segments; and
 (c) trans-splicing at least 2 of the RNA segments together to provide
 at least 1 unencrypted RNA; or alternately
 (d) translating the RNA segments to provide a number of
polypeptide segments and trans-splicing at least 2 of the
polypeptide segments together to provide at least 1 unencrypted
polypeptide;
 (2) parental organisms (II) and progeny organisms (III) produced via
 (I);
 (3) a vector produced via (I);
 (4) a method (III) of unencrypting engineered genetic elements to
 provide at least 1 unencrypted **functional polypeptide**,
 comprising:
 (a) producing an engineered genetic element corresponding to an
 encoded **functional polypeptide**;
 (b) producing a second engineered genetic element corresponding to a
 second encoded **polypeptide** which is **non-functional** in
 the absence of modification by the first **polypeptide**;
 (c) mixing the engineered genetic elements;
 (d) expressing the genetic elements to produce their encoded
polypeptides; and
 (e) modifying the second **polypeptide** with the first to
 produce a **functional** second **polypeptide** and therefore at
 least 1 unencrypted **polypeptide**;
 (5) the first and second genetic elements produced via (III);
 (6) the first and second **polypeptides** produced via (III);

- (7) host cells and parental and progeny organisms produced via (III);
- (8) vectors produced via (III);
- (9) a composition (IV) comprising 1 or more libraries of 2 populations of split genes, the libraries collectively comprising a number of split gene sequence member types in which combinations or subcombinations of the split gene sequence member types collectively correspond to at least 1 complete genetic element;
- (10) a composition (V) comprising 1 or more libraries of 2 populations of enhancer-linked split gene sequences, the libraries collectively comprising a number of enhancer-linked split gene sequence member types, each regulated by a different trans-acting transcription factor, in which combinations or subcombinations of the enhancer-linked split gene sequence member types collectively correspond to at least 1 complete genetic element;
- (11) a method (VI) for recombining non-overlapping gene sequences, comprising:
 - (a) producing a number of non-overlapping gene sequences each corresponding to a different subsequence of a genetic element;
 - (b) producing a number of gap nucleic acid sequences each of which overlaps 2 or more of the non-overlapping gene sequences; and
 - (c) recombining the non-overlapping gene sequences with the gap nucleic acids to provide recombined non-overlapping gene sequences;
- (12) a composition (VII) comprising 1 or more libraries of gap nucleic acids, comprising a number of gap nucleic acid member types, each of which comprises a subsequence identity or is complementary with at least 2 split gene sequence member types; and
- (13) an integrated system (VIII) comprising a computer or computer readable medium, comprising a data set corresponding to a set of character strings corresponding to a set selected from split gene sequences, enhancer-linked split gene sequences, trans-acting transcription factor sequences, engineered genetic elements, non-overlapping gene sequences and gap nucleic acids.

USE - The methods are used for unencrypting trait encrypted gene sequences to provide unencrypted RNA and **proteins**.

Dwg.0/8

FS CPI EPI GMPI

FA AB; DCN

MC CPI: B04-C01; B04-E01; B04-F0100E; **B04-N0400E**; B11-C08E;
B11-C09; **B12-K04E**; B14-S03; C04-C01; C04-E01; C04-F0100E;
C04-N0400E; C11-C08E; C11-C09; **C12-K04E**; C14-S03;
D05-C12; D05-H02; D05-H08; **D05-H09**; D05-H12; D05-H14;
D05-H17; D05-H18

EPI: T01-F15

TECH UPTX: 20001010

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preferred Methods: In (I) the RNA segments comprise trans-splicing introns and the **polypeptide** segments comprise trans-splicing inteins. (I) further comprises selecting at least 1 unencrypted RNA or **polypeptide** for a desired trait or property. At least 1 of the unencrypted **polypeptides** is a full length **protein**. (I) further comprises translating an unencrypted RNA to provide at least 1 additional unencrypted **polypeptide** (at least 1 of which is a full length **protein**), which is/are then selected for a desired trait or property. At least 1 step occurs in vivo or in vitro. At least 1 of the split gene sequences is cDNA. The first set of split gene sequences is produced by mating a first parent organism (with a second group of split gene sequences (Grp2)) with a second parent organism (with a third group of split gene sequences (Grp3)) to produce a progeny organism comprising at least 1 split gene sequence from the second and third groups (Grp2 and Grp3) to provide the first group of split gene sequences (Grp1).

The transcribing step comprises transcribing at least 1 of the Grp1 sequences and at least 1 of the Grp3 sequences to provide the RNA segments. The method further comprises selecting at least 1 progeny organism for a desired trait or property therefore selecting at least 1 unencrypted RNA which is then translated to provide a second unencrypted **polypeptide** (at least 1 of which is a full length **protein**

) which is selected for a desired trait or property.

(II) are either animals, plants bacteria or fungi (especially a yeast), and are preferably plants selected from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersicon*, *Nicotiana*, *Solanum*, *Petunia*, *Capsicum*, *Majorana*, *Cannabis*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Heterocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browallia*, *Lolium*, *Maius*, *Apium*, *Gossypium*, *Vicia*, *Lathyrus*, *Lupinus*, *Pachyrhizus*, *Wisteria* and/or *Stizolobium*. In particular, they are crop plants selected from the genera *Acris*, *Phleum*, *Lactuca*, *Sorghum*, *Setaria*, *Oea*, *Cynna*, *Triticum*, *Secale*, *Avena*, *Hordeum*, *Saccharum*, *Pis*, *Festuca*, *Stenotaphrum*, *Cynodon*, *Cox*, *Glycine*, *Phaseae*, *Glycine*, *Cicer*, *Phaseolus*, *Lens* and/or *Arachis*.

(II) may be corn, rice, cotton, soybean, sorghum, wheat, oats, barley, millet, sunflower, rapeseed, canola, peas, beans, lentils, peanuts, yam beans, cowpeas, velvet peas, clover, alfalfa, lupine, vetch, lotus, sweet clover, wisteria and/or sweet pea.

At least 1 parental organism comprises a number of enhancer-linked split gene sequences, each comprising a subsequence of the genetic element linked to an enhancer sequence. The first parental organism comprises at least 1 trans-acting transcription factor sequence which is unlinked to the first group of enhancer-linked split gene sequences and the second organism comprises a second group of enhancer-linked split gene sequences each comprising a subsequence of the genetic element linked to a second enhancer sequence. The second parental organism also comprises a second trans-acting transcription factor sequence which is unlinked to the second group of enhancer-linked split gene sequences.

The progeny organism comprises at least 1 of the first group of enhancer-linked split gene sequences, at least 1 trans-acting transcription factor sequence, at least 1 of the second group of enhancer-linked split gene sequences, at least 1 of the second trans-acting transcription factor sequences. At least 1 of the first and second enhancer-linked split gene sequences are transcribed to produce the RNA segments (at least 1 of the first group of enhancer-linked split gene sequences is regulated by the second trans-acting transcription factor and at least 1 of the second group of enhancer-linked split gene sequences is regulated by the first trans-acting transcription factor).

The method further comprises selecting the progeny organism for a desired trait to select an unencrypted RNA which is then translated to produce unencrypted **polypeptides** (at least 1 of which is a full length **protein**) which may then be selected for a desired trait.

At least 1 of the enhancer-linked split gene sequences and the trans-acting transcription factors are cDNA. The split gene sequences are preferably subsequences of toxic genes which are spliced together and expressed in the progeny to produce a toxic **polypeptide** which renders the progeny incapable of reproducing if it is male but capable of reproduction if the progeny is female and vice versa. In the male is incapable of reproduction, the female progeny can reproduce to produce hybrid progeny organisms which do not express the toxic gene and vice versa if the female progeny are incapable of reproduction.

The first group of split gene sequences (Grp1) is produced by infecting a host organism comprising Grp2 with a vector comprising Grp3 to produce a transfected organism comprising at least 1 from each of Grp1 and Grp3. At least 1 of the vectors may be a virus.

In (III), the modification of the second non-functional **polypeptide** comprises glycosylation, proteolysis, farnesylation, cholesterol esterification, acetylation, methylation, phosphorylation and/or dephosphorylation. At least 1 step may occur *in vivo* or *in vitro*. The engineered genetic elements may be cDNA.

The first engineered genetic element encodes biotin ligase and the second encodes an engineered biotin dependent glycosylate resistance **polypeptide**. The mixing step comprises mating parental organisms comprising the first or second genetic elements to produce progeny comprising both of the engineered genetic elements. The first and second **proteins** are then expressed by the progeny. The parental organisms

are as described above.

Host organisms comprising the first engineered genetic element are infected with a vector encoding the second to produce a organism comprising both (alternatively, a host comprising the second may be infected with a vector comprising the first genetic element), the genetic elements may then be expressed in progeny.

The vector is preferably a virus.

Method (VI) further comprises:

- (i) selecting the recombinant non-overlapping gene sequences for at least 1 desired property;
- (ii) recombining the recombinant non-overlapping gene sequences; and
- (iii) repeating the selection and recombination steps until a desired recombinant genetic element is obtained.

The non-overlapping gene sequences are derived from a *cr3Bb* gene and the gap nucleic acid sequences are derived from a *cr3Bb*, *cr1C*, and/or a *cr3A* gene. At least 1 step may occur *in vivo* or *in vitro*.

Preferred Compositions: in (IV) at least 3 populations comprise homologous genetic elements.

In (V), a trans-acting transcription factor from 1 population of enhancer-linked split gene sequences regulates the enhancer-linked split gene sequences of the other.

Preferred System: (VIII) further comprises a sequence search and comparison set for searching for specific nucleic acid sequences. It also comprises an automatic sequencer coupled to an output from the computer or computer medium which accepts instructions from the computer or medium which direct sequencing of sequences selected from split gene sequences, enhancer-linked split gene sequences, trans-acting transcription factor sequences, engineered genetic elements, non-overlapping gene sequences and gap nucleic acids. It may also comprise an automatic synthesizer coupled to an output from the computer or computer medium which accepts instructions from the computer or medium which direct synthesis of the set selected from split gene sequences, enhancer-linked split gene sequences, trans-acting transcription factor sequences, engineered genetic elements, non-overlapping gene sequences and gap nucleic acids. (VIII) further comprises 1 or more robotic control elements for incubating, denaturing, hybridizing and elongating a set of recombinant non-overlapping gene sequences and gap nucleic acids and/or a detector for detecting a nucleic acid produced by elongation of the set of recombinant non-overlapping gene sequences and gap nucleic acids, or an encoded product of them.

L123 ANSWER 13 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 2000-548589 [50] WPIX
 DNN N2000-405865 DMC C_000-163688
 TI *Streptococcus pneumoniae fabZ proteins* useful for diagnosing and treating microbial infections.
 DC B04 D16 S03 T01
 IN KONSTANTINIDIS, A; RUSSELL, R B; WARREN, P V
 PA (SMIK) SMITHKLINE BEECHAM CORP
 CYC 19
 PI WO 2000030662 A1 20000602 (200050)* EN Slip A61K035-74
 FW: AT BE CH CY DE DK ES FI FR GR IE IT LU MC NL PT SE
 W: JP
 ADT WO 2000030662 A1 WO 1999-0726435 19991109
 PRAI US 1998-196188 19981119
 IC ICM A61K035-74
 ICS A61K035-01; A61K035-10; C07K016-00; C12N001-00; C12N015-00;
 C12M015-31; C12P021-01; C12Q001-68; G01N033-68; G06F017-30
 AB WO 200030662 A UPAP: 20001019
 NOVELTY - Nucleic acids (I) encoding a **polypeptide** (II) designated *fabZ*, a member of the fatty acid biosynthetic pathway family of **proteins** from *Streptococcus pneumoniae*.
 DETAILED DESCRIPTION - INDEPENDENT CLAIMS are included for the following:
 '1) an isolated polynucleotide (I), comprising either:
 (a) a polynucleotide comprising a sequence encoding a **polypeptide** that has 70-95% identity with a fully defined 140

amino acid sequence (IIa) (given in the specification);

(b) a polynucleotide comprising a sequence that has 70-95% identity to a polynucleotide sequence encoding (IIa) over its entire length;

(c) a isolated polynucleotide comprising a sequence that has 70-95% identity to a defined 423 nucleotide sequence (Ia) given in the specification;

(d) a polynucleotide which encodes (IIa);

(e) a polynucleotide comprising (Ia);

(f) a polynucleotide obtained by screening a library under stringent conditions with a probe comprising (Ia);

(g) an isolated polynucleotide sequence encoding a mature fabZ protein of *S. pneumoniae*; and/or

(h) a sequence complementary to (a)-(g);

(2) a **polypeptide** selected from:

(a) a **peptide** comprising an amino acid sequence having 70-95% identity to (IIa) over its entire length;

(b) a **peptide** comprising (IIa);

(c) a **peptide** that is (IIa); and/or

(d) a **polypeptide** encoded by a recombinant polynucleotide comprising (Ia);

(3) an antibody (III) antigenic for (Ia);

(4) a method (IV) for the treatment of an individual:

(a) in need of enhanced activity or expression of (II), comprising:

(i) administering an agonist of (II); and/or

(ii) providing the individual with (II) to produce the **polypeptide** (and its activity) in vivo; or

(b) in need of inhibited activity of (II), comprising:

(i) administering an antagonist of (II); and/or

(ii) administering a nucleic acid molecule that inhibits the expression of (I); and/or

(iii) administering a **polypeptide** that competes with the **polypeptide** for its ligand, substrate or receptor;

(5) a method (V) for diagnosing or progressing a disease or a susceptibility to a disease in an individual related to expression or activity of (II), comprising:

(a) determining the presence or absence of a mutation in the sequence (IIa) in the **genome** of the individual; and/or

(b) analyzing and quantifying the presence of the **polypeptide** (II) in a sample from the individual;

(6) a method (VI) for screening to identify compounds that activate or inhibit the **function** of (II), comprising:

(a) measuring the binding of a candidate compound (CC) to (II) or to cells or membranes bearing (II) (of a fusion protein of (II)) using a label directly or indirectly associated with the CC;

(b) measuring the binding of a CC to (II) or to cells or membranes bearing (II) (of a fusion protein of (II)) in the presence of a labeled competitor;

(c) testing whether a CC results in a signal generated by activation or inhibition of (II), using detection systems appropriate to the cells or membranes bearing (II);

(d) mixing a CC with a solution containing (II) and measuring the activity of the polypeptide in the mixture, and comparing that activity to a standard;

(e) detecting the effect of a CC on the production of mRNA encoding (II) and (II) in cells using, for example, an enzyme linked immunosorbant assay; and/or

(f) a method comprising:

(i) contacting (II) with the CC and assessing their interactions ;

and

(ii) determining whether the compound interacts with and activates or inhibits the activity of (II) by detecting the presence or absence of a signal generated from the interaction of the compound and polypeptide);

(7) an (anti)agonist (VII) of the expression or activity of (I);

(8) an expression system (VIII) comprising the polynucleotide (I) and capable of expressing (II) when present in a host cell;

(9) a host cell (IX) comprising (VIII) or a membrane expressing (II);

(13) a process (X) for producing a polypeptide comprising culturing (IX);

(14) a process (XI) for producing (IX) or a membrane of (IX) expressing (II), comprising transforming/transfected a cell with (VIII), so that the cell expresses (II);

(15) a computer readable medium (XII), upon which is stored:

(a) (Ia) or sequence(s) comprising (Ia) (and/or other polynucleotide sequences);

(b) (IIa) or sequence(s) comprising (IIa) (and/or other polypeptide sequences);

(c) a data set representing (Ia);

(d) a data set representing polynucleotides encoding (IIa);

(16) a computer based method (XIII) for performing homology identification, comprising providing a sequence comprising (Ia) in a computer readable medium and comparing that sequence to other polypeptides and polynucleotide to identify sequence homology;

(17) a computerized method (XIV) for polynucleotide assembly, comprising providing a nucleic acid comprising (Ia) in a computer readable medium and screening for at least 1 overlapping region between that polynucleotide sequence and a second polynucleotide sequence; and

(18) a polynucleotide (XV) of the formula:

5' X-(R1)m-(R2)-(R3)n-Y 3'

X = H, a metal or a modified nucleotide residue or together with Y defines a covalent bond;

Y = H, a metal or a modified nucleotide residue or together with X defines a covalent bond;

R1 and R3 = any nucleic acid residue or modified nucleic acid residue;

R2 = the nucleotide sequence (Ia); and

m and n = 0-100.

ACTIVITY - Bactericide.

MECHANISM OF ACTION - Vaccine.

No data given.

USE - (I) and (II) may be used in the prevention, treatment and diagnosis of diseases associated with fabZ expression and Streptococcal infection.

(I) or (II) may be administered to treat diseases by rectifying mutations or deletions in a genome that affect the activity of fabZ by expressing inactive proteins or to supplement the production of fabZ polypeptides. Additionally, (I) may be used to produce fabZ, according to standard recombinant DNA methodology, by inserting the nucleic acids into a host cell and culturing the cell to express the protein (either *in vitro* or *in vivo*). Antisense nucleic acid molecules may be administered to down regulate fabZ expression by binding with the cell's own fabZ genes and preventing their expression.

(I) and complementary sequences may be used as probes in diagnostic assays to detect the presence of fabZ in samples.

The polypeptides may be used as antigens in the production of antibodies against fabZ and in assays to identify modulators of fabZ expression and activity. The anti-fabZ antibodies and fabZ antagonists may also be used to down regulate fabZ expression and activity. They may be used to treat *S. pneumoniae* infections.

The anti-fabZ antibodies may also be used as diagnostic agents for detecting the presence of fabZ polypeptides in samples.

Dwg.076

FS CPI EPI
FA AB; DCN
MC CPI: B04-B04C2; B04-B04M; B04-C01G; B04-E03F; B04-E04; B04-E05; B04-E06;
B04-E08; B04-F0100E; B04-G07; B04-N03A0E; B11-A; B11-C07A4; B11-C08E;
B11-C09; B12-K04A4; B12-K04E; B12-K04F;
B14-A01B1; B14-S03; B14-S11B; D05-A01A4; D05-A01B; D05-A02B; D05-C12;
D05-H04; D05-H07; D05-H08; D05-H09; D05-H11; D05-H12A;
D05-H12B; D05-H12E; D05-H14; D05-H17A5; D05-H17A6; D05-H18
EPI: S03-E14H; T01-J05B

TECH UPTX: 20001010

TECHNOLOGY FOCUS - BIOLOGY - Isolation: (I) and (II) were isolated from *S.*

pneumoniae.

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preparation: (I) and (II) encode may be prepared by standard methods.

L123 ANSWER 14 OF 22 WITX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 2000-043909 (49) WITX
 DNN 12000-462327 DDC C2000-161024
 TI Polynucleotides comprising polymorphic variants of a reference sequence for tumor necrosis factor receptor 1 (TNFR1), useful for studying the biological function of TNFR1 and identifying drugs targeting the protein for treating disease(s).
 DC E04 D16 T01
 IN CHEW, A; NANDAALAN, K; SCHULZ, V P; STEPHENS, J C
 FA (CHEW-1) CHEW A; (GENA-1) GENARANCE PHARM INC; (NAND-1) NANDAALAN K;
 (SCHU-1) SCHULZ V P; (STEP-1) STEPHENS J C
 CYC 2000-050436 A1 20000831 (200049)* EN 74P G07H01-62
 FW: AT BE CH CY DE DK EA ES FI FP GB GH GM GR IE IT KE LS LU MC MW NL
 OA PT SB SE SL SS TZ UG SW
 W: AE AL AM AT AJ AS BA BB BG BR BY CA CH CN CP CU CS DE DK DM EE ES
 FI GR GE GH HI HU IL IN IS JP KE KG KP KE LZ LC LK LR LS
 LT LU LV MA MD MG MN MW MX NC NG PL PT EG RU UD SE SG SI SK SL
 TI TM TF TT TZ UA US UG VN YP SA SW
 AU 2000036919 A UPAP: 20000104 (200063); G07H01-62
 ADT WO 2000050436 A1 WO 2000-034666 20000213; AU 2000036919 A AD 1000-36029
 10000223
 FDT AU 2000036919 A Based on WO 2000050436
 PRAI US 1999-131314 133001/23
 IC 13M G07H01-01
 13S A1IN043-04; G13P021-00; G13P021-01; G06F017-00
 AB WO 2000050436 A UPAP: 20001006
 NOVELTY - An isolated polynucleotide (N1), or its complement, comprising a polymorphic variant of a reference sequence for tumor necrosis factor receptor 1 (TNFR1) gene, comprising at least 1 polymorphism selected from a thymine at PS (polymorphic site) 1, PS5, PS6, PS9 and PS10, adenine at PS2, PS4, PS7 and PS11, guanine at PS3, and cytosine at PS6 and PS12, is new.
 DETAILED DESCRIPTION - An isolated polynucleotide (N1) comprising a polymorphic variant of a reference sequence for tumor necrosis factor receptor 1 (TNFR1) gene, comprising a thymine at PS1, PS5, PS6, PS10, adenine at PS2, PS4, PS7 and PS11, guanine at PS3, and cytosine at PS6 and PS12, is new. The reference sequence comprises defined 1139, 1358, or 2254 base pair (bp) sequences (given in the specification).
 INDEPENDENT CLAIMS are also included for the following:
 (1) a recombinant organism transformed or transfected with N1, where the organism expresses a TNFR1 **protein**;
 (2) an isolated polynucleotide (N2) comprising a polymorphic variant of a reference sequence for the TNFR1 cDNA, where the reference sequence comprises a defined 368 bp sequence (given in the specification), comprising a polymorphism selected from guanine at nucleotide 36, thymine at nucleotide 224 or 269, adenine at nucleotide 362, and cytosine at nucleotide 403;
 (3) a recombinant organism transformed or transfected with N2 expressing TNFR1;
 (4) an isolated **polypeptide** (P1) comprising a polymorphic variant of a reference sequence for the TNFR1 **protein**, where the reference sequence comprises a defined 455 amino acid sequence (given in the specification), comprising a leucine at amino acid 75, methionine at amino acid 98, glutamine at amino acid 121 and histidine at amino acid 135;
 (5) an isolated antibody to P1;
 (6) a composition comprising at least 1 genotyping oligonucleotide for detecting a polymorphism in the TNFR1 gene at a polymorphic site selected from PS1-PS12;
 (7) a method (M1) for genotyping the TNFR1 gene of an individual

comprising determining for the two copies of the TNFR1 gene present in the individual, the identity of the nucleotide pair at at least 1 PS selected from P51-P512;

(3) a method (M2) for haplotyping the TNFR1 gene of an individual as in M1 except that the identity of the nucleotide at the PS is for one copy of the TNFR1 gene;

(4) a method (M3) for predicting a haplotype pair for the TNFR1 gene of an individual comprising:

(a) identifying a TNFR1 genotype for the individual at at least 2 polymorphic sites selected from P51-P512;

(b) enumerating all possible haplotype pairs which are consistent with the genotype;

(c) accessing data containing the TNFR1 haplotype pairs determined in a reference population; and

(d) assigning a haplotype pair to the individual that is consistent with the data;

(10) a method (M4) for identifying an association between a trait and at least 1 genotype or haplotype of the TNFR1 gene comprising comparing the frequency of the genotype or haplotype in a reference population, where the genotype or haplotype comprises a nucleotide pair or nucleotide located at P51-P512, where a higher frequency of a genotype or haplotype in the trait population than in the reference population indicates that the trait is associated with the genotype or haplotype;

(11) a computer system for storing and analyzing polymorphism data for the TNFR1 gene comprising:

(a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device; and

(d) a database containing the polymorphism data; and

(12) a method (M5) for screening for drugs targeting TNFR1 isoforms comprising contacting the TNFR1 isoform with a candidate agent and assaying for binding activity, where the reference sequence and polymorphic variants are as described in (4).

ACTIVITY - None given.

MECHANISM OF ACTION - None given.

USE - Determination of the polymorphisms in the TNFR1 gene is useful for studying the biological **function** of TNFR1 as well as for identifying drugs targeting the **protein** for treatment of disorders related to its abnormal expression or **function**.

Dwg.0/5

FS CPI EPI

FA AB; DCM

MC CPI: B04-B03C; B04-C01G; B04-E03D; B04-E03D; B04-E04; B04-E05; B04-F0100E;
B04-G04; B04-K01K0E; B11-C08E; **B12-K04A; B12-K04E**
; **B12-K04F; D05-H09; D05-H11; D05-H12A;**
D05-H11B1; D05-H12D1; D05-H12D5; D05-H14; D05-H17B4

EPI: T01-J

TECH

UPTM: 200001006

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preferred Polynucleotide:N1 comprises a TNFR1 isogene. N1 is DNA and further comprises regulatory elements.

Preferred Organism: The recombinant organism is a transgenic animal.

Preferred Composition: The genotyping oligonucleotide is an allele-specific oligonucleotide that specifically hybridizes to an allele of the TNFR1 gene at a region containing the polymorphic site.

Preferred Method: In M1 and M2 the determining step comprises:

(a) isolating a nucleic acid mixture comprising both copies of the TNFR1 gene;

(b) amplifying a target region containing at least 1 of the polymorphic sites;

(c) hybridizing a genotyping oligonucleotide to one allele of the amplified target region;

(d) performing a nucleic acid template-dependent primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least 2 different terminators of the reaction, where the terminators are complementary to the alternative nucleotides present at the polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended denaturing oligonucleotide.

In M4 the trait is a clinical response to a drug targeting TNFR1.

L123 ANSWER 15 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 1000-431075 [43] WPIX
 CR 1000-43449 [25]; 1000-437367 [41]; 2000-514667 [41]
 DNN 02000-364425 DDC C2000-147612
 TI In silico recombinant nucleic acid preparation by genetic algorithm guided gene synthesis involves providing a number of parental character strings, providing oligonucleotides and elongating them.
 DC B71 D1- TO1
 IN DEL CARBAYRE, S; GUSTAFSSON, C; MINCHILLI, J; PATTEN, P A; SELIFONOV, S A; ZEMMER, W P C; TOBIN, M
 PA (MANY-1) MANYGEN INC
 CYC 01
 P1 WO 200042560 A1 20000726 (2004-5) * EU 123p G06F (19-0) ---
 RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL
 OA PT SD SE SL SE TZ DG SW
 W: AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE DK DM EE ES
 FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS
 LT LU LV MA MD MG MK MN MW NK NJ NZ PL PT RU SD SE SG SI SK SL
 TJ TM TR TT TZ UA US US VN YE SA SW
 AG 1000032101 A 20000801 (2005-4) G06F (19-0) ---
 EP 1002614 A1 20001227 (2001-2) EU G06F (19-0) ---
 R: AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
 ADT WO 1000042560 A1 WO 2000-032101; AU 1000032101 A AU 1000-32101
 00000116; EP 1002614 A1 EP 2000-032101 20000116, WO 2000-032102 20000116
 FDT AU 1000032101 A Based on WO 200042560; EP 1002614 A1 Based on WO 200042560
 PPAI US 1999-416837 20000118; US 1999-116447 10000119; US 1999-118613
 14990265; US 1999-118654 14990265; US 1999-141043 14990243; US
 1999-408392 19990928; US 1999-408393 19990928; US 1999-416870
 19991012; US 1999-416837 19990112; US 1999-416837 20000118
 IC ICM G06F019-00
 AB WO 200042560 A UPAR: 20010110
 NOVELTY - Preparing recombinant nucleic acid (I) from oligonucleotides which correspond to a set of character string subsequences (SCSS) comprising at least two parental character strings (PCS) corresponding to a number of nucleic acids, is new.
 DETAILED DESCRIPTION - Preparing recombinant nucleic acid (I) by aligning for maximum identity a number of parental character strings (PCS) corresponding to a number of nucleic acids, defining a set of character string subsequences (SCSS) comprising at least two of the PCS, providing a set of oligonucleotides corresponding to the SCSS and then annealing and elongating one or more oligonucleotides with polymerase or ligating at least two with ligase to produce (I).
 INDEPENDENT CLAIMS are also included for the following:
 (1) preparing character strings (CS) by providing PCS encoding a polynucleotide or polypeptide, providing a set of oligonucleotide character strings of preselected length that encode a number of single-stranded oligonucleotide sequence comprising sequence fragments of PCS and its complement and creating a set of derivatives of parental sequence comprising sequence variant strings, a set of multiple mutations with one mutation per variant string;
 (2) a library prepared by the above said method;
 (3) facilitating recombination between two or more divergent nucleic acids by aligning PCS corresponding to divergent nucleic acids, identifying regions of sequence identity and regions of sequence diversity, defining a diplomat CS which is intermediate in PCS, synthesising at least a portion of the diplomat sequence to produce a diplomat nucleic acid and recombining a mixture of parental nucleic acid and diplomat nucleic acid;
 (4) a mixture of selected nucleic acids prepared by the above said method;
 (5) generating and recombining nucleic acids by inputting a number of amino acid sequence character strings (ACCS) into a digital system,

reverse translating ASGS in the digital system to a number of nucleic acid character strings which are species codon biased in a selected expression host and with optimized sequence similarity between a number of nucleic acid character strings and synthesizing one or more oligonucleotides from one or more reverse translated nucleic acid sequences;

(6) optimizing activity of a nucleic acid by parameterizing a number of nucleic acids or **proteins** to provide a set of multidimensional datapoints, extrapolating one or more postulated multidimensional datapoint from the set of multidimensional datapoints and converting the postulated multidimensional datapoint to a new CS corresponding to a postulated nucleic acid or **protein**;

(7) providing a library of recombinant nucleic acids which is enriched for a sequence of interest and selecting the library by producing an initial library of at least about 106 recombinant nucleic acids, comprising at least about 105 different non-identical units, hybridizing the library to one or more population of nucleic acids that correspond to one or more subsequences in the different library units;

(8) the enriched library produced by the above said method;

(9) generating a library of biological polymers by generating a diverse population of CS in a computer, which in turn are generated by alteration of pre-existing CS, synthesizing the diverse population of CS in which diverse population comprises the library of biological polymers; and

(10) an integrated system comprising a computer having a first data set comprising a first CS, a second data set comprising a second CS, software for aligning the first and second CSs, software for performing a genetic operation on the first or second CS, an output file comprising a third data set comprising a third CS, the third CS comprising CS subsequences from the first and second CSs, and an oligonucleotide sequence output file comprising a plurality of overlapping oligonucleotide sequences corresponding to third CS.

USE - The method is useful for rapid **evolution** of nucleic acids *in vitro* and *in vivo* and provides for generation of encoded molecules with new and/or improved properties. **Proteins** and nucleic acids of industrial, agricultural and therapeutic importance can be created or improved through DNA shuffling procedures.

ADVANTAGE - Physical access to genes or organisms is not required as sequence information is used for design and selection of oligo. Extensive sequence information is provided and sequences from inaccessible, non culturable organisms can also be used. Sequences from pathogens without actual handling of pathogens and all type sequences including damaged and incomplete genes are amenable to this method. All genetic operators and crossovers can be fully and independently controlled in a reproducible fashion removing human error and variability from physical experiments with DNA manipulations. Sequences with frame-shift mutations are eliminated or fixed. Wild type parents do not contaminate derivative libraries with multiple redundant parental molecules.

Swg.0715

FS CBI EPI
 FA AB; DCN
 MC CBI: B04-E01; B04-E02; B04-E05; B11-B; B11-C09; D06-C07; D06-H10;
 D06-H12A; D06-H12D
 EPI: T01-1

TECH CPTM: 160001907

TECHNOLOGY FOCUS - COMPUTING AND CONTROL - Preferred System: Character strings are aligned in a digital computer or in a web-based system. Optimizing the activity of nucleic acid further involves synthesizing the postulated nucleic acid or **protein**, analyzing principle components of the set of multidimensional data points, shuffling the postulated nucleic acid with an additional nucleic acid and parameterizing the set of nucleic acids or **proteins** by correlating each residue to a matrix of numeric indicators. The matrix is graphically represented as a tetrahedron having an assigned origin at its center with each corner representing a numeral with each residue being positioned at different corner thereby producing the matrix of numeric indicators. Each multidimensional datapoint is correlated by partial least square

projections to latent structure analysis with an output vector to identify a relationship between a matrix of dependent Y variables and a matrix of predictor X variables. Each multidimensional datapoint comprises one or more different parameters that are plotted against each other in n-dimensional hyperspace comprising at least one dimension for each parameter. Population of nucleic acid is fixed to a solid substrate that comprises a column matrix material and a nucleic acid chip. Initial library is produced by recombining one or more homologous nucleic acids or by providing PCS and providing a set of oligonucleotides for elongating them with polymerase.

The integrated system further comprises an oligonucleotide synthesis machine for synthesizing overlapping nucleotides, a number of oligonucleotides encoded by overlapping oligonucleotides and which produce the third nucleic acid encoded by third CS when incubated in one or more cycles of chain extension and a program with an instruction set for applying one or more genetic operator to the first or second CS. The operator is a mutation, a multiplication, a fragmentation of the string or strings, a crossover between one or more strings, a ligation of strings, an elitism calculation, an alignment, a calculation of sequence homology or sequence similarity, a recursive use of one or more genetic operator for **evolution** of CSs, randomness, a deletion mutation, an insertion mutation, and death.

TECHNOLOGY FOCUS - BIOLOGY - Preferred Method: CS comprises one region of similarity when aligned for maximum identity. At least one PCS is an **evolutionary** or artificial intermediate and corresponds to a designed nucleic acid which represents an energy minimized design for an encoded **polypeptide**.

The method comprises applying one or more genetic operator to one or more PCS or one or more CSs. The genetic operator is:

- (1) a mutation, multiplication or fragmentation of the one or more PCS or one or more CS subsequences;
- (2) a crossover between any of the one or more PCS or one or more CS subsequences or an additional CS;
- (3) a ligation of the one or more PCSs or one or more CS subsequence;
- (4) an elitism calculation;
- (5) a calculation or sequence homology or sequence similarity of aligned strings;
- (6) a recursive use of one or more genetic operator for **evolution** of CSs;
- (7) application of a randomness operator to the one or more PCS or the one or more CS subsequences;
- (8) a deletion, an insertion or subtraction mutation of one or more PCS or one or more CS subsequences with an inactive sequence;
- (9) selection of the one or more PCSs or one or more CSs or one or more of CS subsequences with an active sequence;
- (10) or death of the one or more PCSs, or one or more of CS subsequences.

The method further involves selecting a diplomatic sequence comprising an intermediate level of sequence similarity between two or more CS.

The recombinant nucleic acid with a desired property is selected.

Recombinant nucleic acid with one or more homologous nucleic acids or oligonucleotide units for desired trait or property is selected by *in vivo* or *in vitro* selection assay, or by parallel solid phase assay. Recombinant nucleic acid is deconvoluted by sequencing or digesting, sequenced, cloned or synthesized by error-prone assembly polymerase chain reaction (PCR). Preparation of CS further involves reiterative shuffling or selection of library of recombinant nucleic acids.

Generating and recombining nucleic acids further involves fragmentation of one or more elongated nucleic acids, hybridizing the secondary fragments with each other or with a set of primary fragmented nucleic acid that encode amino acid polymer.

Pre-existing CS is altered by recombination. Biological polymers are selected from nucleic acids, **polypeptides** and **peptide**

nucleic acids. Additional library or additional set of CS is filtered by subtracting additional library or additional set of CS with units of biological polymer library which display activity below or above the desired threshold.

Preferred Oligonucleotides: The set of oligonucleotides comprise a number of overlapping nucleotides and SCSS is defined by selecting a length for CS and subdividing at least two of the PCS into segments of selected length. The set of oligonucleotides is provided by synthesizing the oligonucleotides to comprise one or more modified PCS subsequence. The set of single-stranded oligonucleotides that correspond to SCSS are pooled, hybridized and extended as double-stranded nucleic acids which comprise sequences from at least two of the PCS. The double-stranded nucleic acids are denatured to produce a heterogeneous mixture of single-stranded nucleic acids which are then re-hybridized and extended with polymerase. The set of oligonucleotide is synthesized to comprise one or more modified PCS subsequences comprising:

- (1) a replacement, deletion, insertion, inclusion of degenerate sequence at one or more randomly or non-randomly selected positions;
- (2) inclusion of CS from a second PCS subsequence at one or more positions;
- (3) a PCS subsequence biased based upon its frequency in a selected library of nucleic acids; or
- (4) PCS subsequences comprising one or more artificially included sequence motif such as an N-linked glycosylation sequence, an O-linked glycosylation sequence, a protease sensitive sequence, a collagenase sensitive sequence, a RNase-dependent transcriptional termination sequence, an RNA secondary structure sequence that affects the efficiency of transcription, RNA secondary structure sequence that affects the efficiency of translation, a transcriptional enhancer sequence, a transcriptional promoter sequence or a transcriptional silencing sequence.

The oligonucleotide set contains one or more altered or degenerated positions compared to PCS and comprises a chimeric nucleic acid sequence and at least one oligonucleotide unit whose subsequences are separated by a crossover point which is selected by identifying a number of PCS, aligning the substrings to display pairwise identity between the substrings and selecting the point within the aligned sequence as the crossover point. The crossover point is selected randomly or non-randomly by selecting approximately in the middle of one or more identified pairwise identity regions. The method further involves adding one or more oligonucleotide units of the set of oligonucleotides at a concentration higher than additional oligonucleotide units and incubating them with recombinant nucleic acid and polymerase. Recombinant nucleic acid is denatured, contacted with additional nucleic acids produced by cleaving parental nucleic acid from PCS by chemical cleavage, using DNase or restriction endonuclease.

The oligonucleotide set comprises one or more oligonucleotide units of 20-60 nucleotides in length.

Preferred Protein: PCS encodes erythropoietin (EPO), insulin, a peptide hormone, a cytokine, epidermal growth factor, fibroblast growth factor, hepatocyte growth factor, insulin-like growth factor, interferon, interleukins, keratinocyte growth factor, leukemia inhibitory factor, oncostatin M, PD-ECSF (undefined), platelet derived growth factor (PDGF), pleiotropin, SCF (undefined), c-kit ligand, VEGF (undefined), G-CSF (undefined), oncogene, tumor suppressor, steroid hormone receptor, plant hormone, disease resistance gene, herbicide resistance gene, bacterial gene, monooxygenases, protease, nucleic acid or lipase.

L123 ANSWER 16 OF 73 WPIX COPYRIGHT 2001 DEFENT INFORMATION LTD
 AN 2000-481361 [42] WPIX
 CR 2000-203449 [26]; 2000-491075 [41]; 2000-514667 [41]
 DNN H2000-358956 DNC C000-14556
 TI Recombining homologous nucleic acids to produce family shuffle nucleic acids comprises hybridizing and elongating a set of family gene shuffling oligonucleotides and providing a population of recombinant nucleic acids.
 DC B04 D16 T01
 IN BASS, S H; CRAMERI, A; GUSTAFSSON, C; MINSHILL, J; NESS, J E; PATTEN, P A;
 STEMMER, W P C; WELCH, M
 PA (MAXY-N) MAXYGEN INC
 CYC 71

PI WO 2000042561 A2 20000720 (2000421)* EN 74p G06F019-00 <--
 RW: AT BE CH CY DE DM EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL
 OA PT SE SH SL CZ TJ UG ZW
 W: AE AL AM AT AU AS BA BB BG BR BY CA CH CN CR CU CZ DE DK DM EE ES
 FI GR GO GE GH GM HP HU IS IL IN IS JP KE PG KP KR KZ LC LF LS
 LT LU LV MA MD MG MP MN MW MK NO NG NL PT RO SU SI SE SG SI SF SL
 TJ TM TR TT TZ UA UG US UZ YU ZA CW
 AS JOC 000042561 A 20000101 (200054) G06F019-00 <--
 EP 107,010 A2 20000101 (200054) EN G06F019-00 <--
 R: AT BE CH CY DE DM ES FI FR GB GP IE IT LI LU NC NL PT SE
 ADT WO 2000042561 A1 WO 2000-031203 000-0118; AU 2000032102 A AU 2000-32102
 21000111; EP 107,010 A2 EP 107,010-00021 2000 11-, WO 2000-031203 20000118
 FDT AU 2000032102 A Based on WO 200042561; EP 107,010 A2 Based on WO 200042561
 PRA1 US 1999-416337 19991012; US 1999-116447 19991119; US 1999-116513
 19990201; US 1999-116514 19990201; US 1999-116515 19990201; US
 1999-408693 19990313; US 1999-408696 19990328; US 1999-416375
 19991012
 IC ICM G06F019-00
 AB WO 2000042561 A UPAE: 20001027
 NOVELTY - Recombining (R) homologous nucleic acids (I) comprises
 hybridizing and elongating a set of family gene shuffling oligonucleotides
 (II) and providing a population of recombinant nucleic acids.
 DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the
 following:
 (i) introducing nucleic acid family diversity during nucleic acid
 recombination comprises:
 (i) providing a composition comprising at least a set of fragmented
 nucleic acids and a population of (II);
 (ii) recombining (II) with at least one of the fragmented nucleic
 acids;
 (iii) and regenerating a recombinant nucleic acid comprising a
 nucleic acid subsequence corresponding to (II);
 (iv) recombining homologous or **non-homologous**
 nucleic acid sequences having low sequence similarity comprises
 recombining one or more set of fragmented nucleic acids with a set of
 crossover nucleic acids which individually comprise a number of sequence
 diversity domains (SDD) corresponding to a number of SDD from homologous
 or **non-homologous** nucleic acids with low sequence
 similarity and producing a recombinant nucleic acid;
 (v) providing an oligonucleotide set for recombining homologous
 nucleic acids comprises:
 (i) aligning a number of (I) to identify one or more regions of
 sequence heterogeneity; and
 (ii) synthesizing a number of different oligonucleotide units
 corresponding to at least one of the regions of heterogeneity to provide a
 set of oligonucleotides comprising a region of sequence heterogeneity
 corresponding to (I);
 (vi) family shuffling PCR amplicons comprising:
 (i) providing a number of non-homogeneous homologous template nucleic
 acids;
 (ii) providing a number of PCR primers which hybridize to
 non-homogeneous homologous template nucleic acids
 (iii) producing a number of PCR amplicons by PCR amplification; and
 (iv) recombining amplicons to provide recombinant nucleic acid;
 (v) recombining a number of parental nucleic acids comprising
 ligating or elongating a set of number of oligonucleotides comprising a
 number of nucleic acid sequences from parental nucleic acids to produce
 recombinant nucleic acids encoding a full length **protein**;
 (vi) producing a recombinant nucleic acid comprising;
 (i) transducing a population of cells with a set of overlapping (II);
 and
 (ii) permitting recombination to occur between the set of (II) and
 one or more nucleic acids contained within the population of cells to
 provide a population of recombinant nucleic acids within the resulting
 population of recombinant cells;
 (7) a population of recombinant nucleic acids and recombinant cells

produced in (6);

(8) an amplified nucleic acid (III) produced 6;

(9) a cell, a vector, or a virus produced by the above said method;

(10) a composition comprising a library of oligonucleotides comprising a number of oligonucleotide units corresponding to a number of subsequence regions of selected set of homologous target sequence units;

(11) recombining two or more sequences comprising:

(i) aligning two or more nucleic acids to identify regions of identity and regions of diversity;

(ii) providing a non-equimolar set of oligonucleotides corresponding to two or more nucleic acid sequences at least a region of diversity; and

(iii) extending the oligonucleotides with a polymerase to provide a number of recombinant nucleic acids;

(12) making a library of chimeroplasts by providing a number of homologous chimeroplasts each comprising a marker or other region of sequence similarity and at least a region of sequence diversity and producing a library of chimeroplast; and

(13) a library produced in (12).

USE - The method is useful to produce a family of shuffled nucleic acids, to produce recombinant molecules with greater molecular diversities and for generating classical mutagens.

ADVANTAGE - Homologous nucleic acids with low sequence similarity or non-homologous nucleic acids are easily recombined. The fragmented oligonucleotide derived from homologous or non-homologous nucleic acids can hybridize to one or more regions of crossover sites, facilitating recombination.

DESCRIPTION OF DRAWING(S) - The figure shows oligonucleotide-directed *in vivo* shuffling using chimeroplasts.

Dwg.078

FS CPI EPI

FA AB; SCN

MC CPI: B04-E03; B04-E08; B04-F01; B04-F11; B11-C03E3; B11-C03E1; B12-K04F; D05-H12A; D05-H12E; D05-H12F; D05-H14

EPI: T01-J

TECH UPTX: 20000905

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preferred Oligonucleotide: A set of (II) contains overlapping nucleotides and encodes an **evolutionary** intermediate nucleic acid. (II) further comprises a number of oligonucleotide units having consensus region subsequences derived from a number of first round screened nucleic acids or derived from homologous target nucleic acids. (II) further comprises module shuffling oligonucleotide units, each comprising a first subsequence from a first sequence module and a second subsequence from a second sequence module. (II) also comprises a number of codon-varied oligonucleotides and a number of oligonucleotide units comprising at least 3-10 units in approximately equimolar amounts or in non-equimolar amounts. Recombinant nucleic acid is selected from one or more desired trait or property, and encodes a full length **protein**. The set of overlapping (II) are chimeroplasts.

The library of oligonucleotide comprises at least 10-50 or more oligonucleotide units in non-equimolar amounts and each have a sequence identical to set of homologous target sequences comprising at least one variant subsequence. A number of subsequent regions comprises non-overlapping sequence regions on the selected set of homologous target sequences. The composition comprising a library of oligonucleotides further comprises polymerase, thermostable DNA polymerase, nucleic acid synthesis reagent, buffer, salt, magnesium, and one or more nucleic acids comprising a number of homologous target sequence units.

Preferred Method: (R) involves elongating using polymerase or ligase. The method further comprises denaturing the population of recombinant nucleic acids, reannealing the denatured recombinant nucleic acids, extending or ligating the resulting reannealed nucleic acids and optionally selecting one or more recombinant nucleic acids for desired property. (R) further involves recombining the reannealed recombinant nucleic acids. Numerous populations of recombinant nucleic acids can be screened and determined for a desired property to provide first round screened nucleic acids by hybridizing a second set of (II), derived from first round screened

nucleic acids and elongating the second set of (II) to provide a population of further recombinant nucleic acid sequences. The second set of (II) comprises numerous oligonucleotides comprising subsequences corresponding to at least one region of diversity. The composition comprising at least one fragmented nucleic acid and a population of (II) is provided by aligning homologous nucleic acid sequences to select conserved regions of sequence identity and regions of sequence diversity, synthesizing a number of (II) corresponding to sequences comprising regions of sequence diversity, providing a full length nucleic acid identical to or homologous with homologous nucleic acids, fragmenting the full length nucleic acid by cleavage with DNase or by partial chain elongation, and mixing the resulting set of nucleic acid fragments with a number of (II) to provide a composition comprising fragmented nucleic acids and a population of (II). The method further comprises selecting at least a second full length nucleic acid and cleaving it to provide a second set of nucleic acid fragments which is also mixed with population of (II). (II) are added to the composition by aligning homologous nucleic acid sequences, selecting at least one conserved region of sequence identity and a number of regions of sequence diversity comprising a number of domains and synthesizing (II) corresponding to domains of sequence diversity. (II) causes domain switching with the fragmented nucleic acid in the regenerated recombinant nucleic acid as compared to homologous nucleic acid sequences and encodes one or more domains of sequence diversity. Fragmented nucleic acid is obtained by cleaving a cloned nucleic acid or selecting a nucleic acid sequence and by synthesizing oligo nucleotide fragments corresponding to selected nucleic acid sequence. A number of oligonucleotide sequence are synthesized serially or in parallel. Homologous nucleic acid sequences are aligned in a system comprising computer with software for sequence alignment or by manual alignment. Oligonucleotide unit are selected by aligning a number of homologous target sequences, determining regions of identity and diversity, synthesizing oligonucleotides to encode at least a portion of region of identity, a region of diversity or both the regions and selecting them. Sequence diversity domains corresponds to adjacent sequence regions of homologous nucleic acids when aligned. Producing chimeroplast library further includes transducing a population of cells with the library and detecting the recombination of marker or other regions of similarity with one or more nucleic acids to identify the recombinant homologous chimeroplast and thus to identify active homologous chimeroplast. Active homologous chimeroplasts are recombined to produce a library which are then transduced with a second population of cells and identifying additional active homologous chimeroplasts.

L123 ANSWER 17 OF 73 WPIX COPYRIGHT 2001 DEFVENT INFORMATION LTD
 AN 2000-482840 [42] WPIX
 CR 1999-372624 [32]
 DNC C/000-145334
 TI Novel methods for selecting target sites for, and production of, zinc finger **proteins**, useful for controlling expression of target genes, e.g. for inhibiting oncogenes or treating sickle cell anemia.
 DC B04 D16
 IN CASE, C C; COX, I G H; EISENBERG, S P; JAMESON, A; REBAF, E J; COX, G N
 PA (SANG-HEI) SANGAMO BIOSCIENCES INC
 CYC 00
 PI WO 2000042219 A1 20000720 (200042) * EN 82p Cl. Q01-68
 FW: AT BE CH CY DE DK EA ES FI FR GB GH GM GF IE IT HE LU MW NL
 OA PT SD SE SL SG TC UG EW
 W: AE AL AM AT AU AS PA BB BG BR BY CA CH CN CF CU CS DE DK DM EE ES
 FI FR GD GE GH GM HF HU ID IL IN IS JP KE KG PP KP KC LC LF LS
 LT LU LV MA MD MG MK MN MW NX NO NE PL PT RU SE SF SG SI SK SL
 TJ TM TF TT TE GA UC US YE VN YU SA EW
 GB 1948425 A 20001004 (200051) G01F017-50 1--
 AU 2000027220 A 20000301 (200044) Cl. Q01-68
 ACT WO 2000042219 A1 WO 2000-03334 20000106; GB 1948425 A GB 2000-651
 20000112; AU 2000027220 A AU 2000-07220 20000106
 FDT AY 2000027220 A Based on WO 200042219

IRAI US 1999-229007 19991112
 IC ICM C12Q011-68; G06F017-50
 AB WO 200042119 A UPAb: 20000005

NOVELTY - Selecting a target site (TS) within a nucleic acid (I) to be targeted by a zinc finger **protein** (ZFP) by detecting a specific 10-base motif (A), is new.

DETAILED DESCRIPTION - Novel method for selecting a target site (TS) within a nucleic acid (I) to be targeted by a zinc finger **protein** (ZFP) by detecting a specific 10-base motif of formula (A).

5'-NNx aNy bNzic-3' (A)

each of (x, a), (y, b) and (z, c) = (N, N) or (G, K) but at least 1 must be (N, K);

N = any nucleotide;

K = G or T.

INDEPENDENT CLAIMS are also included for the following:

- (a) selecting a TS by:
 - (i) identifying potential TSs comprising three contiguous triplets;
 - (ii) determining subscores for each combination of triplets and triplet positions by applying a correspondence regime;
 - (iii) combining subscores for all three triplets to give a score for the potential TS;
 - (iv) repeating the procedure for at least one other potential TS; and
 - (v) outputting at least 1 potential TS together with its score;
- (b) producing ZFP comprising constructing a database of many ZFPs where the nucleic acid sequences for each ZFP comprise at least 3 triplets bound specifically by the individual fingers of the ZFP, and in the same order (3' to 5') as the fingers are arranged (N to C) in the **protein**, identifying fingers that specifically bind to each triplet, and outputting the (sub)designations of the relevant ZFPs;
- (c) a computer program for selecting a TS, comprising code for providing a polynucleotide sequence, selecting a potential TS within the sequence, calculating a score for the potential TS from a combination of subscores for 3 triplets as in (a), repeating the selection and calculation steps at least once, and providing an output of the TS with its score;
- (d) a system for selecting a TS, comprising a memory, a system bus, a processor operatively disposed to provide or receive a polynucleotide sequence, select a potential TS and calculate a score for the TS as in (c);

(e) a computer program for designing a ZFP, comprising code for a database of 3 fingered ZFPs with subdesignations for each finger and a corresponding nucleic acid for each ZFP, providing a TS, identifying ZFPs in the database that specifically bind to the TS and outputting the designations and subdesignations of the ZFPs;

(f) a system for selecting a TS, comprising a memory, a system bus, a processor operatively disposed to provide or a database of ZFP designations and output ZFP designations and subdesignations as in (e).

ACTIVITY - Antibacterial; antiviral; cytostatic; neuroprotectant; antianemic.

MECHANISM OF ACTION - Modulation of gene expression by the binding of a ZFP.

USE - Selection of TS is used to design ZFP that bind to preselected targets. ZFP bind to DNA and can modulate (inhibit or activate) the expression of a wide range of genes. Typical of many potential applications, of ZFP or the nucleic acids that encode them, are: inhibition of bacterial or viral genes, oncogenes or the *apoE* gene (implicated in Alzheimer's disease); inducing expression of fetal hemoglobin (for treating sickle cell anemia); and in plants to increase resistance to diseases or herbicides, or to increase oleic acid synthesis at the expense of linolic or linolenic acids. ZFP can also be used diagnostically, e.g. to detect variant, disease-related alleles; to quantify copy numbers of a gene; to detect pathogenic microbes, and in analysis of phenotype and **function** of gene expression.

ADVANTAGE - ZFP can be controlled by small molecules, allowing the adjustment of the degree of repression or activation produced by ZFPs and, in transgenic animals, makes switching on a ZFP at a late stage in

embryonic development possible, so that effects can be studied in the adult. Nucleic acids encoding a ZFP can be introduced at any site (homologous recombination is not required) and because ZFPs are trans-dominant, only 1 chromosomal copy need be present (functional knockouts can be produced without backcrossing).

Ewg.07/3

FS CPI

FA AB; DGN

MC CPI: **B04-N04**; B11-C03; **B12-K04A**; **B12-K04D**; **B12-K04E**; B14-A01; B14-K02; B14-F03; B14-H01B; B14-H01A4; B14-N16; D05-C13; D05-H04; D05-H05; D05-H06; D05-H09; L05-H17A5; L05-H17B6

TECH UPTM: 20000005

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preferred Target Sites: In (A) at least 2, preferably all 3 of (x, a), (y, b) and (z, c) are NK and TS may include two (A), preferably separated by at most 1 bases. Preferably, one segment is 5'-N8G-3' and the other is 5'-KNxGyBNGc-3', with no intervening bases. TS may occur in a coding region or within, or outside, a regulatory region.

Preferred Process: Many different potential TSs are selected from a target gene. The ZFP of (A) has 3 fingers, 1 specific for each of NK, any and NNG. Preferably each finger is selected and designed independently from a database of ZFP that contains designations of ZFP, subdesignations of individual fingers and nucleic acid sequences bound by the fingers.

Alternatively, a finger is selected by screening variants of a zinc finger binding **protein** for specific binding to TS, thus identifying a variant that binds to TS. In the method of (a), at least 1 potential TS, with the highest score, is the output and the subscores are combined as their products. The correspondence regime comprises 74 triplets and 192 subscores. Subscores are determined by assigning different scores for a subset of triplets and corresponding positions for which there is an existing ZFP with a finger specific for the triplet from the same position as in the correspondence regime, a subset for which an existing finger specifically binds to the triplet from a different position, and a subset for which there is no known finger that specifically binds. The specification includes a table showing the correspondence regime. The subscore of at least 1 triplet may be combined with a context parameter (CP) to give a scaled subscore, preferably when the target site comprises the sequence 5'-NNGK-3'. A CP may also be combined with the score for a potential target to give a scaled score, particularly where the potential TS contains motif (A), and different CP are used depending on how many (x, a), (y, b) and (z, c) are (G, K) are included. A similar process to (b) may be used to produce a two-triplet target site and 2-finger ZFP. Once ZFP and nucleic acid sequences encoding them have been designed, the **proteins** can be produced by standard recombinant DNA methods.

L123 ANSWER 18 OF 73 WPIX COPYRIGHT 2001 DEFENTWENT INFORMATION LTD

AN 2000-442645 [38] WPIX

DNN N2000-330228 DNC C2000-134729

TI 3-D structure of sterile alpha motif domain used as model for determining 3-D structures of additional native or mutated SAM domain with unknown structure and structures of co-crystals of SAM domain with modulators.

DC B04 T01

IN SICHERI, F; STAPLETON, B

PA (MOUN) MOUNT SINAI HOSPITAL

CYC 3u

PI WO 2000037500 A1 20000629 (200036)* EN 73p C07F014-705 <--
FW: AT BE CH CY DE DK EA ES FI FR GB GH GM GP IE IT KE LS LU MC MW NL

OA PT SD SE SL SS TZ US EW

W: AE AL AM AT AU AS BA BB BG BR BY CA CH CN CP JU CZ DE DK DM EE ES
FI GB GD GE GH GM HF HV ID IL IN IS JP KE KG KP KE KC LC LK LR LS
LT LU LV MA MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL
TJ TM TR TT TZ UA UG US UZ VN YU SA EW

AU 2000017642 A 20000712 (200046) C07F014-705 <--

ADT WO 2000037500 A1 WO 1999-CA1209 19991217; AU 2000-17642 A AU 2000-17642
19991217

FDT AU 200017642 A Based on WO 20037500

PRAI U.S. 143-112923 19981213

IC I.M C07K014-705

IDS A61K038-17; A61P025-00; C07K014-47; G06F017-50

AB WO 20037500 A UPAB: 20000811

NOVELTY - A purified three dimensional structure of a **polypeptide** (I) corresponding to one or more sterile alpha motif (SAM) domains.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(1) a purified crystalline form (CF) of (I);

(2) forming a CF of (I);

(3) a potential modulator (II) of **function** of a SAM domain of an Eph (Erythropoietin producing human hepatocellular carcinomas cell line) receptor identified using CF; and

(4) a **peptide** (III) which mediates SAM domain

function comprises VVSV, SAVVSV, FAVNV, FSAVVNV, FSAVVSVD, VVSVNLWL, FNTVDS, FNTVDEWL, TCFNTVDEWL, TCFNTV, YTSFNTV, RSEV, RSEVLG, RSEVLCVD, VPFRSEV or VIFRSEVLGW or has a formula I or II:

(I) X1-X2-X3-X4-X5-X6 ;

(II) X7-X8-X9-X10-X11-X12-X13-X14-X15-X16.

X = defined in DEFINITIONS: Full Definitions field

ACTIVITY - Cytostatic; immunosuppressive; vasoactive; neuroprotective; neurotropic; antiparkinsonian; cerebroprotective; antiarthritic; antiatherosclerotic.

MECHANISM OF ACTION - Eph (Erythropoietin producing human hepatocellular carcinomas cell line) receptor SAM domain modulators (claimed).

USE - The structural coordinates of a SAM domain or CF of (I) can be used as a model for determining the three dimensional structures of **polypeptides** with SAM domains of unknown structure. They are also used for identifying (II) a potential modulator of a SAM domain of a Eph (Erythropoietin producing human hepatocellular carcinomas cell line) receptor **function** which involves docking a computer representation of a structure of a compound with the computer representation of a structure of one or more SAM domains of an Eph receptor. (II) identified by the above method or a CF, can be used for treating a disease associated with inappropriate activity of SAM domain of an Eph receptor in cellular organism, such that a SAM domain **function** is activated or inhibitor to treat a cell proliferative disease such as cancer, angiogenesis, atherosclerosis, arthritis and diseases associated with the nervous system (claimed). The knowledge of the three dimensional structure of a SAM domain, in particular the EphA4 SAM domain enables us to identify homologues. The structure coordinates of a SAM domain or CF of (I) can be used for identifying additional native or mutated SAM domains with unknown structure, as well as the structures of co-crystals of SAM domains with compounds such as modulators. The structure coordinates and models of a SAM domain three dimensional can also be used to determine solution-based structures of native or mutant SAM domains. The structural coordinates of a SAM domain structure may be applied to nuclear magnetic resonance (NMR) data to determine the three dimensional structures of **polypeptides**. The rational design and identification of modulators of SAM domains can be accomplished by utilizing the atomic structural coordinates that define SAM domain's three dimensional structure. The new **peptides** (III) may be used to identify lead compounds for drug development. A comparison of the structure of **peptides** similar in sequence but differing in the biological activities they elicit in target molecules can provide information about the structure-activity relationship (SAR) of the target. Information obtained from the examination of SAR can be used to design either modified **peptides** or other small molecules or lead compounds which can be tested for predicted properties as related to the target molecule. The new **peptides** may be used to prepare antibodies. Antibodies and labeled antibodies specific for the new **peptides** may be used to screen for the **proteins** containing SAM domains. They may be used in the treatment and diagnosis of disorders associated with aberrant T cell signaling and to modulate

telomere function. In particular, they may be useful in methods for therapy of cellular senescence and immortalization controlled by telomere length and telomerase activity and selective immunosuppressants (e.g. in organ transplantation). They may also be useful in the treatment of cancer. The new three dimensional SAM domain structure, new **peptides** and modulators may be used to modulate the biological activity of an Eph receptor or Eph ligand in a cell and in particular regulating a pathway in a cell regulated by the ligand or the receptor, particularly those pathways involved in neuronal development, axonal migration, pathfinding and regeneration. The new three dimensional SAM domain structure **peptides** and modulators identified using the methods will be useful in pharmaceuticals to modulate axonogenesis, nerve cell interactions and regeneration, to treat conditions such as neurodegenerative diseases and conditions involving trauma and injury to the nervous system, for example Alzheimer's disease, Parkinson's disease, Huntington's disease, demyelinating diseases such as **multiple sclerosis**, amyotrophic lateral sclerosis, bacterial and viral infection of the nervous system, deficiency diseases such as Weil-Marcus disease and nutritional polyneuropathy, progressive supranuclear palsy, Shy Drager's syndrome, multi-system degeneration and olive pontocerebellar atrophy, peripheral nerve damage, trauma and ischemia resulting from stroke. The pharmaceutical composition may be used to stimulate or inhibit neuronal development, regeneration and axonal migration associated with neurodegenerative conditions and involving trauma and injury to the nervous system.

Dwg.0/3

FS CFI: EPI
 FA AR; SCN
 MC CFI: B04-C01; B14-A01; B14-A02; P11-709; B14-F01; B14-H01; B14-J01A5;
 B14-J01A4; B14-N16; B14-S01
 EPI: T01-519

TECH UPTX: 2000C0811

TECHNOLOGY FOCUS - ORGANIC CHEMISTRY - Preparation: The CF of (I) is formed by mixing a volume of a SAM domain with a reservoir solution; and incubating the mixture obtained in previous step over the reservoir solution in a closed container under conditions suitable for crystallization (claimed).

Preferred Crystalline Form: CF has dimensions of about $a=b=77.14\pm 0.3$ angstroms, $c=24.3\pm 0.04$ angstrom and has a **protein** coordinate data as given in the patent specification.

Preferred Method: (M1) involves docking a computer representation of a compound from a computer database with computer representation of a selected site on a three dimensional structure of a SAM domain of an Eph receptor or CF, determining a conformation of the complex with a favorable geometric fit and favorable complementary interactions and then identifying compounds that best fit the selected site as potential modulators of SAM domain **function**. Alternately, the method involves modifying computer representation of a compound complexed with a selected site on a three dimensional structure of a SAM domain of an Eph receptor or CF by deleting or adding a chemical groups, determining a conformation of the complex with a favorable geometric fit and favorable complementary interactions and then identifying a compound that best fits the selected site as potential modulators of SAM domain **function**. The method can also be performed by selecting a computer representation of a compound from a computer database with computer representation of a selected site on a three dimensional structure of a SAM domain of an Eph receptor or CF, and then searching for molecules in a database that are similar to the compound using a computer searching program, or replacing portions of the compound with similar chemical structures from a database using a compound building computer program.

Preparation: (III) is prepared by standard solid phase **peptide** synthesis.

Preferred 3D Structure: The 3D structure of SAM domain is of the SAM domain of an Eph receptor preferably EphA, complexed with one or more compounds and comprising one or more heavy metal atoms.

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preparation: (III) is also prepared by standard recombinant techniques.

L123 ANSWER 19 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 2000-387560 [33] WPIX
 DNN N2000-386184 DNO C2000-11758.
 TI New DnaB **polypeptide** from *Streptococcus pneumoniae*, useful, e.g., in vaccines, for diagnosis of infections, and for identifying antibacterial agents.
 DC B04 D16 T01
 IN CHAMBER, A F; HOLMES, D J; INGRAM, K A; JAWORSKI, C D; LENNOX, A L; MAY, S W; MAGNUSSA, M; PAY, J; WANG, M; WARFEN, R L
 PA (MMB) SMITHKLINE BEECHAM CORP
 CYC 13
 PI WO 2000128626 A1 20000513 (199833)* EN 53p A01NC07-18
 RW: AT BE CH CY DE IE ES FI FR GB GF IE IT LU MC NL PT SE
 W: JP
 ADT WO 2000128626 A1 WO 1994-08-23 19991111
 PFAI US 1998-191879 19981113
 IC ICM A01NC07-18
 ICS A01NC061-00; C07H021-02; C07H021-04; C07K001-00;
 C07K016-00; C12N15-00; C12P011-06; G06F017-00
 AB WO 2000128626 A1 WPAE: 20000712
 NOVELTY - Isolated DnaB **polypeptide** (I) that is at least 70% identical with a 450 residue amino acid sequence, fully defined in the specification, over the entire length of it, comprises, or is, the 450 residue sequence, or is encoded by a recombinant polynucleotide comprising a 1953 base pair sequence, fully defined in the specification.
 DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:
 (1) polynucleotide (II) that
 (a) encodes the sequence, at least 70% identical to the 450 residue sequence;
 (b) is at least 70% identical with a sequence encoding the 450 residue sequence;
 (c) is at least 70% identical with the 1953 base pair sequence over the entire 301-1651 nucleotide (nt) segment of it;
 (d) encodes the 450 residue sequence;
 (e) is the 301-1651 nt segment of the 1953 base pair sequence;
 (f) is isolated by screening a library, under stringent conditions, with the 1953 base pair sequence, or a fragment of it;
 (g) encodes a mature **polypeptide** expressed by the DnaB gene of *Streptococcus pneumoniae*; or
 (h) is a complement of (a)-(g);
 (2) antibody (Ab) immunospecific for (I);
 (3) diagnosis or prognosis of disease, or susceptibility, related to expression or activity of (I), comprising determining the presence or absence of a mutation in the nucleotide sequence encoding the **polypeptide** in the **genome** of the individual, and analyzing for the presence or amount of the **polypeptide** expression in a sample derived from the individual;
 (4) screening methods for identifying compounds (A) that activate or inhibit **function** of (I), comprising:
 (a) measuring the binding of a candidate compound to the **polypeptide**, or to cells or membranes bearing the **polypeptide** or a fusion **protein** of it, using a label directly or indirectly associated with the candidate compound;
 (b) measuring the binding of a candidate compound to the **polypeptide** or to the cells or membranes bearing the **polypeptide** or a fusion **protein** of it, in the presence of a labeled competitor;
 (c) testing if the candidate compound results in a signal generated by activation or inhibition of the **polypeptide**;
 (d) mixing a candidate compound with a solution containing (I), measuring the activity of the **polypeptide**, and comparing it to a

standard;

(i) detecting the effect of a candidate compound on the production of mRNA encoding the **polypeptide**, and the **polypeptide** in cells, using e.g. enzyme linked immunosorbent assay (ELISA); or

(ii) contacting a composition comprising the **polypeptide** with the compound to be screened to assess the interaction, the interaction being associated with a second component capable of providing a detectable signal in response to the interaction of the **polypeptide** and compound, and determining if interaction occurs;

(i) agonists and antagonists of the expression or activity of (I);

(ii) expression system comprising (II), present in a host cell;

(iii) host cell, or its membrane, that contains the system of (ii) and expresses (I);

(iv) production of (I) by culturing cells of (ii);

(v) production of cells of (ii), or its membranes, by transformation or transfection;

(vi) computer-readable medium containing at least the 450 residue sequence and/or the 1353 base pair sequence;

(vii) computer-based method of homology identification, based on the 1953 base pair sequence;

(viii) computer-based method of polynucleotide assembly based on identification of an overlap between the 1953 base pair sequence, and a second nucleic acid sequence; and

(ix) polynucleotides of formula (IIa) X-(E1)m-R2-(E2)m-Y (IIa).

X and Y = hydrogen, metal, modified nucleotide or together form a covalent bond;

each E1 and E2 = optionally modified nucleotide;

m = 0-100;

R2 = optionally modified 1353 base pair sequence.

ACTIVITY - Antibacterial.

MECHANISM OF ACTION - Inhibition of DnaB, probably a replicative helicase, which is essential for growth and/or survival of *S. pneumoniae*.

USE - The 450 residue polypeptide, the product of the DnaB gene of *Streptococcus pneumoniae*, is used to screen for specific agonists and antagonists, potential therapeutic agents, to raise specific antibodies (Ab), in vaccines, and in rational drug design. Ab are useful as diagnostic immunoassay reagents and as therapeutic antagonists. Nucleic acids (I) that encode (I), or fragments, are used for recombinant production of (I), and as probes and primers to isolate homologous full-length or genomic clones, for diagnosis, prognosis, staging and typing infections, including detection of genomic mutations, and for chromosome identification or mapping. (II) can also be used in genetic immunization, and as antisense inhibitors. The therapeutic agents have bacteriostatic/bactericidal activity and are used to treat or prevent infections, especially those caused by *S. pneumoniae*, but also *Helicobacter pylori* infections and associated disorders, also for treatment of in-dwelling devices and wounds to prevent bacterial adhesion.

Dwg.070

FS CPI EPI

FA AB; D01

MC CPI: B04-C01G; B04-E01F; B04-E05; B04-E06; B04-E05; B04-F10B4; B04-G01; B04-N03A0E; B11-C07A; B11-C07A4; B11-C08E5; B12-K04A; B12-K04E; B12-K04F; B14-A01; B14-A01B2; B14-S11B; D05-C12; D05-H09; D05-H11; D05-H12A; D05-H14; D05-H17A6

EPI: T01-J

TECH UPTM: 20000712

TECHNOLOGY FOCUS - BIOLOGY - Preferred treatment: When increased expression/activity of (I) is required, an agonist of (I) or a (I)-encoding nucleic acid is administered, and when inhibition is required, an antagonist of (I), nucleic acid that inhibits (I) expression or a **polypeptide** that competes with (I) for ligand, substrate or receptor is administered.

Preferred diagnosis: This involves detecting a mutation in the **genomic** sequence encoding (I) or analysis for presence or amount of (I).

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preparation: (I) is isolated from *Streptococcus pneumoniae* 31/0993 (NCIMB 40300) by standard methods of cloning and screening (no more details); once isolated, (I) can be expressed in any conventional vector/host system, optionally as a fusion protein.

1123 ANSWER TO OF 7.3 WPIX DOP/RIGHT 2001 DERWENT INFORMATION LTD
 AN 2000-187059 (33) WPIX
 DNN NC000-290183 DHC C:000-117585
 TI New pseudouridine synthase from *Streptococcus pneumoniae*, useful, e.g. in vaccines, for diagnosis of infections, and for identifying antibacterial agents.
 DC B-4 D10 T01
 IN P.WELL, D.J.
 EA (SMI) SMITHKLINE BEECHAM CORP
 CYC 1
 PI WO 199901414 A1 20000101 (00063) EN 56p A01N037-18
 RW: AT BE CH CY DE DK ES FI FR GB GE IE IT LU MC NL PT SE
 W: JP
 ADT WO 00023819 A1 WO 1999-US26433 19991109
 PRAI US 1999-190824 19981111
 IC ICM A01N037-18
 ICS A01N01-00; C07H011-02; C07H011-04; C07K001-00;
 C07K016-00; C12N015-00; C12P021-06; G06F017-00
 AB WO 2000-08119 A UPAE: 10000712
 NOVELTY - Isolated **polypeptide** (I) from *Streptococcus pneumoniae* is a pseudouridine synthase.
 DETAILED DESCRIPTION - Isolated **polypeptide** (I) is:
 (a) at least 70, 80, 90 or 95% identical over the entire length, to defined amino acid (aa) sequence (II) of 292 aa, given in the specification;
 (b) an isolated **polypeptide** comprising (II);
 (c) (II); or
 (d) is encoded by a recombinant polynucleotide comprising a 1061 bp sequence (III) given in the specification.
 INDEPENDENT CLAIMS are also included for the following:
 (1) an isolated polynucleotide (IV) which:
 (a) is an isolated polynucleotide encoding a **polypeptide** at least 70, 80, 90 or 95% identical over the entire length to (II);
 (b) comprises at least 70, 80, 90 or 95% identity over the entire length to a polynucleotide encoding (II);
 (c) comprises a nucleotide sequence with at least 70, 80, 90 or 95% identity over the entire length of nucleotides 97-973 of (III);
 (d) comprises a nucleotide sequence encoding (II);
 (e) is nucleotides 97-973 of (III);
 (f) is obtained by screening a library, under stringent conditions, with (III) or a fragment of (III);
 (g) encodes a mature **polypeptide** expressed by the YfiI (pseudouridine synthase) gene of *Streptococcus pneumoniae*; or
 (h) is a complement of (a)-(g);
 (i) antibody (Ab) immunospecific for (I);
 (j) methods of treating conditions requiring increased or reduced activity or expression of (I);
 (4) diagnosis or prognosis of disease (or susceptibility) related to expression or activity of (I);
 (5) screening methods for identifying compounds (A) that activate or inhibit **function** of (I);
 (6) agonists and antagonists of the expression or activity of (I);
 (7) expression system comprising (IV), present in a host cell;
 (8) host cell, or its membrane, that contains the system of (7) and expresses (I);
 (9) production of (I) by culturing cells of (8);
 (10) production of cells of (8), or its membranes, by transformation or transfection with (IV);
 (11) a host cell produced by the method of (10);
 (12) computer-readable medium containing at least sequences (II)

and/or (III);

(13) computer-based method of homology identification, based on sequence (III);

(14) computer-based method of polynucleotide assembly based on identification of an overlap between (III) and a second nucleic acid sequence; and

(15) polynucleotides of formula (IIa).

X-(R1)m-R2-(R3)n-Y (IIa)

X and Y = hydrogen, metal, modified nucleotide or together form a covalent bond;

each R1 and R3 = optionally modified nucleotide;

m and n = 0-1000;

R2 = region 55-473 of (I).

ACTIVITY - Antibacterial.

No biological data is given.

MECHANISM OF ACTION - Inhibition of pseudouridine synthase which is essential for growth and/or survival of *S. pneumoniae*.

USE - Sequence (II), the product of the YfiI (pseudouridine synthase) gene of *Streptococcus pneumoniae*, is used to screen for specific agonists and antagonists (potential therapeutic agents), to raise specific antibodies (Ab), in vaccines and in rational drug design. Ab are useful as diagnostic immunoassay reagents and as therapeutic antagonists. Nucleic acids (III) that encode (I), or fragments, are used for recombinant preparation of (I), as probes and primers to isolate homologous full-length or genomic clones, for diagnosis, prognosis, staging and typing infections (including detection of genomic mutations), and for chromosome identification or mapping, in genetic immunization; and as antisense inhibitors. The therapeutic agents have bacteriostatic/bactericidal activity and are used to treat or prevent infections, especially those caused by *S. pneumoniae*, but also *Helicobacter pylori* infections and associated disorders, also for treatment of in-dwelling devices and wounds to prevent bacterial adhesion.

Dwg.676

FS CPI EPI

FA AB; DCN

MC CPI: B04-B03E; B04-B04; B04-F0100E; B04-G01; B04-L05; B11-C03;

B12-K04A4; B12-K04F; B14-A01A; B14-A01B2;

D05-H09; D05-H11; D05-H12A; D05-H12E; D05-H14; D05-H17A3

EPI: T01-3

TECH UETM: 200000712

TECHNOLOGY FOCUS - BIOLOGY - Preferred Treatment: When increased expression/activity of (I) is required, an agonist of (I) or a (I)-encoding nucleic acid is administered, and when inhibition is required, an antagonist of (I), nucleic acid that inhibits (I) expression or a **polypeptide** that competes with (I) for ligand, substrate or receptor is administered.

Preferred Diagnosis: This involves detecting a mutation in the **genomic** sequence encoding (I) or analysis for presence or amount of (I).

Preferred Screening Method: Method comprises:

(i) measuring binding of a test compound (TC) to (I), or cells or membranes carrying (I), or its fusion **protein**, from a label (in)directly associated with TC;

(ii) as (i) but in presence of labeled competitor;

(iii) determining if TC produces a signal generated by activation or inhibition of (I);

(iv) incubating TC with (I), then measuring activity of (I);

(v) determining the effect of TC on production of (I) or mRNA that encodes (I);

(vi) incubating (I) and TC and determining any interaction, from a second component that produces a detectable signal in response to this interaction.

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preparation: (III) is isolated from *Streptococcus pneumoniae* 0100993 (NCIMB 40800) by standard methods of cloning and screening (no more details). Once isolated, (III) can be

expressed in any conventional vector/host system, optionally as a fusion **protein**.

L123 ANSWEPC 21 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 2000-239656 [29] WPIX
 DNN H.1000-294974 IEC C2000-103109
 TI Engineering **proteins** comprising searching a 3D computer database for framework **proteins** which can then be modified, useful for the design and engineering of cytokine mimetics.
 DC B04 516 T01
 IN ANDREWS, P E; DOOLEY, M J; SMYTHE, M L
 FA (AU) UNIV QUEENSLAND
 CYC --
 PI W 2000011411 A1 2000023474 (200037) EN 78p C07K014-61 --
 FW: AT BE CH CY DE DK EA ES FI FR GE GH GM GR JE IT KE LS LU MC MW NL
 OA PT SD SE SL SJ TS CG SW
 W: AE AL AM AT AU AZ BA BR EG FR PT CA CH CN CP CU CZ DE DK IM EE ES
 FI GB GD GE GH GM HF HU ID IL IN IS JP KE KG KP KR FZ LC LK LR LS
 LT LU LV MA MB MG MK MN MW MX NO NZ PL PT RO FU SD SE SG SI SK SL
 TG TM TR TT TS UA 'UG US US VN YU SA ZW
 AU 2000011411 A 2000023474 (200037) C07K014-61 --
 ADT WO 2000023474 A1 WO 1999-AU014 19991021; AU 2000011411 A AU 2000-11411
 19991021

FDT AU 2000011411 A Based on WO 2000023474
 PFAI AU 1999-606 19991021

IC ICM C07K014-61
 ICS C07K014-53; C07K014-54; G06F017-30

AB WO 2000023474 A UPAT: 20000017
 NOVELTY - A method for **protein** engineering comprising searching a 3D computer database for framework **proteins** with desired structural features, is new.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are included for the following:

- (1) a method of **protein** engineering comprising:
 - (a) creating a computer database with a number of entries, each entry corresponding to a description of a location and orientation in 3D space of side chains of amino acid residues of a framework **protein**;
 - (b) creating a query corresponding to a description of a location and orientation in 3D space of the respective side chains of at least 2 amino acid residues of a sample **protein**; and
 - (c) searching the database with the query to identify hits, where at least 1 of the hits corresponds to a respective framework **protein** which has structural similarity to the sample **protein**;
- (2) a modified framework **protein** produced by the method of (1) which is a cytokine mimetic;
- (3) an engineered **protein** comprising at most 70 amino acid residues and 2-11 disulfide bonds of a framework **protein**, together with at least 2 amino acid residues of another **protein** which are non-contiguous in primary sequence;
- (4) an engineered **protein** having an amino acid sequence selected from SCY01, SCY02, SCY03, EEF01, EEF02, EEF03 and VIB01;
- (5) a computer program for searching a **protein** database which comprises a number of entries, each entry corresponding to a distance matrix representation of at least 2 C alpha -C beta vectors, comprising:
 - (a) comparing a query with each database entry; and
 - (b) identifying hits by clique detection, where a hit is defined according to a minimum number of C alpha -C beta vector matches between the query and each entry; and
- (6) a computer program which filters the hits identified at step (b) of (5).

USE - The method is useful for identifying **proteins** suitable for **protein** engineering, particularly for the design and engineering of cytokine mimetics.

ADVANTAGE - The method is advantageous since a modified framework **protein** may display one or more desired characteristics, such as a

function similar to, or inhibitory of the sample **protein**.

Dug.0/14

ES CPI EPI

FA AB; DGN

MC CPI: B04-E08; B04-H02D; B04-H03G; B04-H04A; B04-J05; B04-N02; B11-C08;
B12-K04E; D05-H09

EPI: T01-J05B

TECH UPTK: 20000617

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preferred Method: The framework **protein** identified as a hit has a greater stability than the sample **protein**. The method further comprises modifying an amino acid sequence of the framework **protein** which corresponds to a hit, by substituting at least 1 amino acid residue with at least 1 amino acid residue of the sample **protein** to create a modified framework **protein**. At least two of the amino acid residues of the sample **protein** which substitute amino acid residues of the framework **protein** are non-identical in primary sequence. The modified framework **protein** has increased structural similarity to the sample **protein**. The modified framework **protein** is capable of exhibiting a **function** which is either similar to, or inhibitory of, a **function** of the sample **protein**.

The location and orientation of each amino acid side-chain of the framework **protein** and sample **protein** is simplified as a C-alpha-C-beta vector. The C-alpha-C-beta vector is in the form of a distance matrix representation. At step (a) the hits are ranked according to structural similarity with the sample **protein**. Searching at step (c) includes identification of the hits by clique detection and filtering the hits identified at step (a). The framework **protein** is a small cysteine-rich **protein** capable of internal disulfide bond formation. The small cysteine-rich **protein** comprises at most 70 amino acids, having 1-11 disulfide bonds. The sample **protein** is a cytokine.

Preferred **Protein**: The engineered **protein** exhibits a **function** either similar to, or inhibitory of a cytokine. The cytokine is selected from growth hormone (GH), interleukin (IL)-4, IL-6 and granulocyte-colony stimulating factor (G-CSF).

Preferred Computer Program: The computer program of (5) is a VECTRIX program, preferably a POSTVEC program.

L123 ANSWER 32 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 2000-303449 [26] WPIX

CR 2000-462662 [41]; 2000-491075 [41]; 2000-514667 [41]

DNN NJ000-226724 DNG C2000-09_051

TI Novel methods for recombinant codon-altered libraries of nucleic acids used to produce new **proteins** and new vectors with reduced rates of reversion to wild type.

DC B04 C06 D16 J04 T01

IN LIU, L; FATTEN, P A; STEMMER, W P C

PA (MAXY-N) MAXYGEN INC

CYC 87

PI WO 2000016306 A1 20000406 (200026) * EN 92p C12N015-10

RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LC LU MC NW NL
OA PT SD SE SL SG TG UG SW

W: AF AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB
GD GE GH GM HR HU ID IL IN IS CP KE EG PP KR KZ LC LE LS LT LU
LV MD MG MN MW MX NO NC PL PT RO RU SD SE SG SI SK SL TJ TM TR
TT UA UG US VI YU SA SW

AU 2000011490 A 20000417 (200035) C12N015-10

ADT WO 2000016306 A2 WO 1999-102153; 19990323; AU 20000111930
19990323

FDT AU 20000111930 A Based on WO 200016306

PRAI US 1999-141049 19990624; US 1999-102362 199609.9; US 1999-117729
19990129; US 1999-110615 19990229

IC ICM C12N015-10

ICS A61K039-21; B01J013-00; C07B061-00; C12N005-10; C12N007-04;
C12N015-11; G06F019-00

ICA C12N015-12; C12N015-43; C12N015-55

AB WO 200018306 A UPAB: 10000421

NOVELTY - A method of making codon altered nucleic acids (NA), comprising providing a NA sequence (NA1), which encodes a **polypeptide** (P1), providing codon altered NA sequences, each encoding P1 or a modified form of it, and recombining the codon altered NA sequences to produce a target codon altered NA which encodes a second **protein**, is new.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

- (1) a library of codon altered conservatively, or non-conservatively, modified NAs produced by the novel method;
- (2) the target codon altered NA produced by the novel method;
- (3) the cell, vector or virus comprising a target codon altered NA;
- (4) an attenuated virus, produced by recombining a target codon altered NA with a portion of the viral **genome**, which produces an immune response upon injection of the virus in a mammal;
- (5) a viral vector produced by recombining a target codon altered NA with a portion of the viral **genome**, where optionally the vector requires trans complementation for replication and has a reduced rate of reversion to a replicative form, compared to a wild type vector;
- (6) a method of making a library of codon altered NAs, comprising
 - (a) selecting a NA1 which encodes P1; and
 - (b) making codon altered NA sequences, comprising the library, each encoding P1 or a modified form of it;
- (7) a codon altered library made by the method of (6), which comprises at least 2, 5, 10, or 100 codon altered NAs;
- (8) a composition comprising codon altered NAs each of which encode P1 or a modified form of it;
- (9) a library of codon-altered NAs, comprising codon altered NAs derived from homologous NAs;
- (10) a kit comprising the library of (9), and a container and instructions for recombining two or more members of the library; and
- (11) an integrated system, comprising a computer or computer readable medium comprising a database having at least 2 artificial homologous codon altered NA sequence strings, and a user interface allowing an user to selectively view sequence strings in the database.

ACTIVITY - None given.

MECHANISM OF ACTION - Vaccine.

USE - The methods of the invention are used for recombining codon-altered libraries of nucleic acids to produce new **proteins**, which have improvements in a desirable characteristic. Target nucleic acids include those coding for therapeutic **proteins** such as EPO, insulin, **peptide** hormones, growth factors, cytokines, interferons, interleukins, leukemia inhibitory factor, and oncostatin M, as well as transcription and expression activators and **proteins** from infectious organisms for use as vaccines. The methods can also be used to produce attenuated viruses which have reduced rates of reversion to wild type.

DESCRIPTION OF DRAWING(S) - The figure is a schematic diagram of the human erythropoietin (EPO) wobble sequence space.

Dwg.4/19

FS CPI EPI

FA AB; GI; DCH

MC CPI: B04-E02; B04-E03; B04-E05; B04-E08; B04-F0100E; B04-F1100E;

B04-N04; B11-C06; B11-C06E; B12-K04; B14-S11;

C04-E02; C04-E03; C04-E05; C04-E08; C04-F0100E; C04-F1100E;

C04-N04; C11-C06; C11-C06E; C12-K04; C14-S11;

D05-C07; D05-H07; D05-H09; D05-H12B2; D05-H12E; D05-H12F;

D05-H14; J04-B01

EPI: T01-J05B4P; T01-S03

TECH UPTX: 20000531

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preferred Method: At least one of the codon altered NA sequences does not hybridize to NA1 under stringent conditions. The method further comprises shuffling a NA comprising a subsequence of NA1, or a substantially identical variant, with one or more of the codon altered NAs or with the target codon altered NA. The method

further comprises screening the second **protein** (P2) for a structural or **functional** property, and comparing this property to that of P1. P2 preferably has a structural or **functional** property equivalent to or superior to P1, or P1 and P2 are homologous. The codon altered NAs comprise a library of codon altered NAs, or a library of codon altered conservatively modified NAs. The codon altered NAs is derived from forms of NAl, or comprises at least 3 codon altered NAs. The codon altered NA sequences comprise one or more of: codon usage divergence for each of the codon altered NAs of 5%, 7%, or 10% or more as compared to NAl, maximal codon usage divergence for each of the codon altered NAs as compared to NAl, non-overlapping non-conservative substitutions in each of the codon altered NAs as compared to NAl, lack of high stringency hybridization between one or more codon altered NAs and NAl, and modification of the codons to provide one or more different hydrophobic side residues for an encoded **polypeptide** as compared to the first **polypeptide**. The percent identity between P2 and P1 is lower than the percent identity between two of the codon altered NAs. Each of the codon altered NA sequences comprises at least two differences when compared to NAl. The method further comprises introducing the target codon altered NA into a cell, vector or virus. Preferred Virus: The virus of (5) comprises viral elements from a lentivirus, a herpes virus, and an adeno-associated virus. Preferred System: The system of (1) further comprises an automated oligonucleotide synthesizer operably linked to the computer or computer readable medium, which synthesizer is programmed to synthesize one or more oligonucleotides comprising one or more subsequences of one or more of the artificial homologous codon altered NAs. Preferred library: The library of (4) comprises at least 2, 5, 10 or preferably 100 codon altered NAs.

TECHNOLOGY FOCUS - BIOLOGY - Preferred Polynucleotide: NAl encodes erythropoietin alpha (EPO), granulocyte-colony stimulating factor (G-CSF), a viral envelope **protein**, a cytokine, and a phosphatase.

L123 ANSWER 23 OF 73 WPIX COPYRIGHT 2001 DEFVENT INFORMATION LTD
 AN 2000-256855 (22) WPIX
 DNN N2000-190969 DNC C2000-078429
 TI Novel glycoprotease **polypeptides** and polynucleotides isolated from *Staphylococcus aureus*, used to screen for antibacterial compounds and to diagnose diseases.
 DC B04 D16 S03 **T01**
 IN KOSMATKA, A L; PALMER, L M; TRAINI, C M; WARREN, R L
 PA (SMIK) SMITHKLINE BEECHAM CORP
 CYC 19
 PI WO 2000013694 A1 20000316 (200022)* EN 59p A61K035-74
 RW: AT BE CH CY DE DK ES FI FR GB GR IE IT LU MT NL PT SE
 W: JP
 ADT WO 2000013694 A1 WO 1999-US19073 19990823
 PPAI US 1998-149624 19980908
 IC ICM A61K035-74
 ICS A61K038-04; **C07K016-00**; C12N001-00; C12N015-00; C12N015-31;
 C12P021-00; C12Q001-68; G01N033-65; **G06F017-30**;
G06F159-00
 AB WO 2000013694 A UPAB: 20000508
 NOVELTY - An isolated glycoprotease (gp) **polypeptide** (I) selected from a **polypeptide** having at least 70%, 80%, 90% or 95% identity to a 341 residue amino acid sequence, over its entire length, a **polypeptide** comprising the 341 residue sequence, and a **polypeptide** encoded by a recombinant polynucleotide comprising the 1026 base pair (bp) sequence, is new. All sequences fully defined in the specification.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(1) an isolated polynucleotide (II) selected from a polynucleotide having 70%, 80%, 90% or 95% identity to a polynucleotide encoding (I), a polynucleotide having at least 70%, 80%, 90% or 95% identity to a 1026 bp

sequence, fully defined in the specification, over its entire length, a polynucleotide which has the 1026 bp sequence, a polynucleotide obtained by screening a library under stringent hybridization conditions with a probe having the 1026 bp sequence, or a fragment of it, a polynucleotide encoding a mature **polypeptide** expressed by the *gcp* gene of *Staphylococcus aureus*, and a polynucleotide sequence complementary to the them;

- (2) an antibody antigenic to or immunospecific for (I);
- (3) a method of treating an individual in need of enhanced activity or expression of (I), comprising:
 - (a) administering to the individual an agonist to (I); or
 - (b) providing to the individual (I) in a form to effect production of the activity of (I) *in vivo*;
- (4) a method for treating an individual having need to inhibit activity or expression of (I), comprising:
 - (a) administering an antagonist to (I);
 - (b) administering a nucleic acid molecule that inhibits the expression (I); or
 - (c) administering a **polypeptide** that competes with (I) for its ligand, substrate or receptor;
- (5) a method for screening to identify compounds that activate or that inhibit the **function** of (I), comprising one of the following:
 - (a) measuring the binding of a candidate compound to (I) or to the cells or membranes bearing (I) or a fusion **protein** of it, using a label directly or indirectly associated with the candidate compound;
 - (b) measuring the binding of a candidate compound to (I) or to the cells or membranes bearing (I) or a fusion **protein** of it, in the presence of a labelled competitor;
 - (c) testing if the candidate compound results in a signal generated by activation or inhibition of the **polypeptide** using detection systems appropriate to the cells or cell membranes bearing (I);
 - (d) mixing a candidate compound with a solution containing (I), measuring activity of the **polypeptide** in the mixture, and comparing the activity to a standard;
 - (e) detecting the effect of a candidate compound on the production of mRNA encoding (I) and (I) in cells, using e.g. enzyme linked immunosorbant assay (ELISA); or
 - (f) contacting a composition comprising (I) with the compound to be screened under conditions permitting interaction between the compound and the **polypeptide** to assess the interaction of a compound, the interaction being associated with a second component capable of providing a detectable signal in response to the interaction of the **polypeptide** with the compound, and determining if the compound interacts with and activates or inhibits an activity of (I) by detecting the presence or absence of a signal generated from the interaction of the compound with (I);
 - (6) an agonist or an antagonist of the activity or expression of (I);
 - (7) an expression system comprising a polynucleotide capable of producing (I) when the system is in a compatible host cell;
 - (8) a host cell comprising the expression system of (8), or a membrane of the cell expressing (I);
 - (9) a process for producing (I), comprising culturing a host cell of (8);
 - (10) a process for producing a host cell comprising the expression system of (8) or a membrane of the cell expressing (I), comprising transforming or transfecting a cell with an expression system comprising a polynucleotide capable of producing (I) when present in a host cell;
 - (11) a host cell produced by the process of (11) or a membrane of the cell expressing (I);
 - (12) a computer readable medium on which is stored a member selected from:
 - (a) a polynucleotide comprising the 1026 bp;
 - (b) a **polypeptide** comprising the 341 amino acid sequence;
 - (c) a set of polynucleotide sequences, at least one which comprises the 1026 bp sequence;

- (d) a set of **polypeptide** sequences, at least one of which comprises the 341 amino acid sequence;
- (e) a data set representing a polynucleotide sequence comprising the 1026 bp sequence; and
- (f) a data set representing a polynucleotide sequence encoding a polypeptide sequence comprising the 341 amino acid sequence;
- (13) a computer based method for performing homology identification, comprising providing a polynucleotide sequence comprising the 1026 bp sequence in a computer readable medium, and comparing the polynucleotide sequence to at least one polynucleotide or polypeptide sequence to identify homology; and
- (14) a computer based method for polynucleotide assembly, comprising providing a polynucleotide sequence comprising the 1026 bp sequence, in a computer readable medium, and screening for at least one overlapping region between the polynucleotide sequence and a second polynucleotide sequence.

ACTIVITY - Antibacterial; cytostatic; antiulcer; antiarthritic; vulnerary; immunosuppressive.

Mechanism of Action - The protein is a glycoprotease.

Use - The polynucleotides and polypeptides are used as research reagents for the discovery of treatments and diagnostics for diseases, particularly human diseases. The polynucleotides may be used as a source of hybridization probes, for screening genetic mutations, and serotype, and for organism chromosome identification. The polypeptides can be used to produce antibiotics, in vaccine formulations, to identify agonists and antagonists, and to screen for antibacterial drugs. These are used to prevent, inhibit or treat diseases, particularly of *Helicobacter pylori* infections, such as gastrointestinal carcinoma, gastric ulcers, and gastritis. Diseases diagnosed, treated and prevented by the products include osteomyelitis, septic arthritis, septic thrombophlebitis, acute bacterial endocarditis, *Staphylococcal* food poisoning, scalded skin syndrome, and toxic shock syndrome. The polypeptides can also be used to treat wounds and body implants to prevent bacterial adhesion and infection.

ADVANTAGE - The frequency of *Staphylococcal* infections has risen dramatically, and isolated *S. aureus* strains are increasingly resistant to standard antibiotics. The glycoprotease (gcp) products can be used screen for new antibacterial compounds that may target these resistant bacteria.

Dwg.0/0

FS CPI EPI

FA AB; DCN

MC CPI: B04-C01G; B04-E02E; B04-E08; B04-F010E; B04-G03; B04-L05C0E; B11-C07A4; B11-C08E5; **B12-K04A; B12-K04F;**

B14-A01; B14-A01A; B14-C03; B14-E08; B14-E10B; B14-G02; B14-N17A;

B14-N17B; B14-S06; D05-C03C; D05-H01; D05-H07; D05-H08;

D05-H09; D05-H11; D05-H12; D05-H12E; D05-H14; D05-H17A3;

D05-H18

EPI: S03-E14H4; S03-E14H5; T01-J05B

L123 ANSWER 24 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 2000-J04637 [18] WPIX

DNN N2000-0-152209 DNC C2000-043041

TI Database system for storing biomolecular sequence information in manner which permits sequences to be catalogued and searched according to one or more **protein function** hierarchies.

DC B04 D16 T01

IN AKERBLOM, I E; ALTUS, C M; AU-YOUNG, J; HILLMAN, J L; KLINGLER, T M; MASLYN, T J; RUSSO, F; SEILHAMER, J J

FA (INCY-N) INCYTE PHARM INC

CYC 1

PI US 6013659 A 20000208 (J00013;*) 39p G06F007-08 <--

ADT US 6013659 A Provisional US 1997-28284 19961010, Provisional US 1996-32563 19961112, US 1997-812290 19970316

PRAI US 1997-812290 19970306; US 1996-28284 19961010; US 1996-32563 19961112

IC ICM G06F007-08

AB US 6027659 A UPAB: 20000412

NOVELTY - At least some of the sequences in the database are grouped into a first hierarchy of **protein function** categories specifying biological **functions of proteins** corresponding to the sequences. The first hierarchy includes first and second sets of categories specifying **functions** at a cellular level and at a level above the cellular level respectively. A user interface allows a user to selectively view the information.

DETAILED DESCRIPTION - The sequences may include nucleic acid sequences and amino acid sequences. Records include descriptive information from an external database to which the biomolecular sequences are correlated. The external database may be GenBank. The second set of categories specifies biological **functions** at an organism or tissue level. At least some of the biomolecular sequences are grouped into a second hierarchy of **protein function** categories. These specify either the molecular or enzymatic **functions of proteins** corresponding to the sequences. The interface allows the user to selectively view information regarding a subset of the sequences which is grouped in selected categories from the first and second hierarchies, or sequence records associated with a category selected from the first hierarchy, or categories from the first hierarchy that are associated with a selected sequence record. Searches may be made using both the presence of keywords and the absence of anti-keywords.

INDEPENDENT CLAIM is included for a method and computer readable medium for using a computer to present information of biomolecular sequence records.

USE - The system stores and retrieves biological information. It may be used to catalogue animal sequences, e.g. human, primate, rodent, amphibian, insect, etc., plant sequences, and microbial sequences. The system may be used to establish a DNA profile for a given tissue and to evaluate changes in gene expression caused by disease progression, pharmacological treatment, aging, etc.

ADVANTAGE - The system allows sequence information to be catalogued and searched rapidly according to one or more **protein function** hierarchies.

DESCRIPTION OF DRAWING(S) - The drawing shows a block diagram of client-server Intranet for providing database services.

Dwg.2A/6

FS CPI EPI

FA AP; GI; DCN

MC CPI: B04-B03C; B04-C01; B04-E01; **B04-N04**; B11-C08; B11-C08E4;B11-C09; **B12-K04A3**; **B12-K04E**; D05-H02;

D05-H09; D05-H12; D05-H18A; D05-J

EPI: T01-E01A; T01-J05B1B; T01-J05B4M; **T01-S03**

L123 ANSWER 25 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 2000-182726 [16] WPIX

CR 1999-312973 [26]

DNN 00000-134704 DMC C0000-057323

TI Determining **function of protein or polypeptide** domains, useful for drug discovery for **proteins** involved in neoplastic, metabolic, neuro-degenerative, cardiovascular, psychiatric, inflammatory and infectious disorders.

DC B04 D16 **T01**

IN ANDERSON, S; HUANG, T; MONTELIONE, G

PA (FUT) UNIV RUTGERS STATE NEW JERSEY

CYC 86

PI WO 200005414 A1 20000209 (100016)* EN 71p C122001-68
 FW: AT BE CH CY DE DK EA ES FI FF GB GH GM GF IE IT KE LS LU MC MW NL
 SA PT SD SE SL SG US SW

W: AE AL AM AT AU AC BA BB BG BP BY CA CH CN CU CL DE DE EE ES FI GB
 GD GE GH GM HF HU ID IL IN IS JP KE FG KE KE LC LF LR LS LT LU
 LV MD MG MN MW MK NO NG PL PT RO RU SD SE SG SI SF SL TJ TM TR
 TT UA VG US US VN YU SA SW

AU 9951155 A 20000214 (200029) C122001-68

ADT WP 2000005414 A1 WO 1999-US16417 19990721; AU 9951155 A AU 1999-51155

19980721

FD/T AU 9451155 A Based on WO 200005414

PRAI US 1993-181601 19931229; US 1993-93641 19980721

IC ICM C12Q001-68

ICS C07K013-00; C12P017-34; G06F017-00

AB WO 200005414 A UPAB: 20000617

NOVELTY - A method for determining a biochemical **function** of a **protein** or **polypeptide** domain comprises identifying a stable **polypeptide** domain and comparing 3-dimensional (3D) structure determined by nuclear magnetic resonance (NMR) to a known structure, is new.

DETAILED DESCRIPTION - A novel high-throughput method for determining a biochemical **function** of a **protein** or **polypeptide** domain of unknown **function** comprises:

(a) identifying a putative **polypeptide** domain that properly folds into a stable **polypeptide** domain, the stable **polypeptide** having a defined 3D structure;

(b) determining 3D structure of the stable **polypeptide** domain from an automated analysis of NMR spectrometer spectra of the **polypeptide** domain, where the automated analysis is conducted by a NOESYAssign process;

(c) comparing the determined 3D structure of the stable **polypeptide** domain to known 3D structures in a **protein** data bank, where the comparison identifies known structures within the **protein** data bank that are homologous to the determined 3D structure; and

(d) correlating a biochemical **function** corresponding to the identified homologous structure to a biochemical **function** for the stable **polypeptide** domain.

INDEPENDENT CLAIMS are also included for the following:

(1) an integrated system for rapid determination of a biochemical **function** of a **protein** or **protein** domain of unknown **function** comprising:

(a) a first computer algorithm capable of parsing the target polynucleotide (tPN) into at least one putative domain encoding region;

(b) a designated lab for expressing the putative domain;

(c) a NMR spectrometer for determining individual spin resonances of amino acids of the putative domain;

(d) a data collection device capable of collecting NMR spectral data, where the data collection device is operatively coupled to the NMR spectrometer;

(e) at least one computer;

(f) a second computer algorithm capable of assigning individual spin resonances to individual amino acids of a **polypeptide**;

(g) a third computer algorithm capable of determining tertiary structure of a **polypeptide**, where the **polypeptide** has had resonances assigned to individual amino acids of the **polypeptide**;

(h) a database, where stored within the database is information about the structure and **function** of known **proteins** and determined **proteins**; and

(i) a fourth computer algorithm capable of determining 3D structure homology between the determined 3D structure of a **polypeptide** of unknown **function** to 3D structure of a **protein** of known **function**, where the **protein** of known structure is stored within the **protein** database, where the fourth computer algorithm determines the structure by an automated NOESYAssign process; and

(2) a high-throughput method for determining a biochemical **function** of a **polypeptide** of unknown **function** encoded by a tPN comprising:

(a) identifying at least one putative **polypeptide** domain encoding region of the tPN (parsing);

(b) expressing the putative **polypeptide** domain;

(c) determining if the expressed putative **polypeptide** domain forms a stable **polypeptide** domain with a defined 3D structure (trapping);

- (d) determining the 3D structure of the stable **polypeptide** domain by an automated NMRAssign process;
- (e) comparing the determined 3D structure of the stable **polypeptide** domain to known 3D structures in a **Protein** Data Bank to determine whether any such known structures are homologous to the determined structure; and
- (f) correlating a biochemical **function** corresponding to the homologous structure to a biochemical **function** for the stable **polypeptide** domain.

NMR - The methods can be used for elucidating the **function** of **proteins** and **protein** domains, particularly for drug discovery purposes. They can be used for domains from **proteins** genetically implicated in neoplastic, metabolic, neurodegenerative, cardiovascular, psychiatric and inflammatory disorders, such as Alzheimer's beta peptide precursor **protein** (APP), or domains from **proteins** associated with infectious agents, e.g. bacteria, fungi and viruses.

Dwg.0/23

FS CPI EPI
 FA AB; DCN
 MC CPI: B04-N02; B11-C04A; B12-K04A; D05-H09; D05-H12;
 D05-H14; D05-H17
 EPI: T01-J

TECH UPTK: 20000350
 TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preferred method: The novel method further comprises the prestep of parsing a target polynucleotide into at least one putative **polypeptide** domain. The parsing is performed by a first computer algorithm selected from a computer algorithm capable of determining exon phase boundaries of a polynucleotide and a computer algorithm capable of determining interdomain boundaries encoded in a polynucleotide. The method further comprises a computer algorithm which compares the putative **polypeptide** domain sequence with known domain sequences stored in a database. The NMR spectra are analyzed by a second computer algorithm which automatically assigns resonance assignments to the **polypeptide** sequence. The identification step comprises measuring a time course of amide hydrogen-deuterium exchange. Prior to the second step the **polypeptide** is optimally solubilized, comprising preparing an array of microdialysis buttons, each containing at least 1micro-l of an approximately 1mM solution of the **polypeptide** domain, dialyzing the members of the array against a different dialysis buffer, analyzing the dialyzed buttons to determine if the **polypeptide** domain has remained soluble, and selecting the domain having optimum solubility characteristics for NMR spectroscopy. The comparison of the 3D structures is performed by a third computer algorithm, preferably DALI, CATH or VAST, capable of determining 3D structure homology between the determined 3D structure and a member of the **protein** data bank (PDB).

L123 ANSWER 26 OF 73 WPIX COPYRIGHT 2001 DEWEY INFORMATION LTD
 AN 2000-136797 [12] WPIX
 DNN N2000-102313 DNC C2000-641*90
 TI An efficient, accurate and rapid computer database for estimating **protein functions** e.g. enzymatic activity, for **polypeptides** obtained from gene sequence translation.
 DC B04 D16 T01
 IN IMAMURA, M; ITAI, A; ITAI, R; TOMIOKA, N
 PA (MEIDI-N) INST MEDICINAL MOLECULAR DESIGN INC
 CYC 89
 PI WO 9962004 A1 19991202 (200011)* JA 26p G06F017-30 --
 PW: AT BE CH CY DE DK FA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL
 OA PT SD SE SC SG SW
 W: AL AM AT AU AC BA BB BG BR BY CA CH CN CU CS DE DK EE ES FI GP GE
 GH GM GW HU ID IL IS KE KG KE KZ LC LF LM LT LU LV MD NG MK MN
 MW MK NO NZ PL PT RU SD SE SG SI SM SL TJ TM TR TT UA WG US UZ
 VN YU SW
 AU 9874513 A 19991213 (200012) G06F017-30 --

ADT WO 9962004 A1 WO 1998-JP2302 19980526; AU 9874513 A AU 1998-74513
19980526, WO 1998-JP2302 19980526

FDT AU 9874513 A Based on W. 9962004

PRAI WO 1998-JP2302 19980526

IC ICM G06F017-30

AB WO 9962004 A UPAB: 20000120

NOVELTY - A database containing information on the amino acid (aa) sequences of **proteins** of which 1 or more biological **functions** are known is new.

DETAILED DESCRIPTION - The database also contains additional information on the score of importance of each aa residue in the whole aa sequence in determining the known biological **functions**.

An INDEPENDENT CLAIM is also included for a method of preparing an **alignment** between aa sequences contained in the database and those of the unknown **polypeptide**. This is represented as the homology amongst various sites, such being identified as having a high score of importance in determining potential biological **functions**.

USE - The method is to enable an efficient estimation of the biological **functions**, particularly enzymatic and signaling activities, of (un)known **polypeptides** from their nucleotide sequences. Suitable **proteins** can then be isolated and purified by various means. This could be of considerable use in a biological and medical context.

ADVANTAGE - The computerized procedure is efficient, fast and accurate.

Dwg.0/2

FS CPI EPI

FA AB; SCH

MC CPI: B04-L095; B04-L05A; B04-L05C; B04-N04; P11-C08;
B12-K04E; D05-H09

RI P: P11-C08

TECH UPTK: 20000303

TECHNOLOGY FOCUS - INSTRUMENTATION AND TESTING - Preferred Database: The information contained in the database is particularly constructed from aa sequences of **proteins** with known biological **functions**, particularly information relating to the stereostructures of the applicable **proteins**. The database is incorporated into media with memory for use.

Preferred Method: A search procedure can be operated with the database to find homology between a **protein** in the database and unknown **protein**. A high correlation can be obtained by using sequence groups which contain 2 or more continuous aa residues of high importance score in exhibiting the biological **functions**. A final determination of homology and possible identity is obtained by the **alignment** of a single **protein** in the database and the target **protein**.

L123 ANSWER 27 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 2000-047350 (08) WPIX

DNN H2000-075212 DMC C10009-028260

TI Optimization of agents to regenerate biosensor surfaces.

DC B04 D16 J04 S03 T01

IN ANDERSSON, E; HÄMEELÄINEN, M; MALMQVIST, M; RÖDS, H

PA (BIAC-H) BIACORE AB

CYC 21

PI WO 9963333 A1 19981120 (200001) * EN 15lp G01N027-327

FW: AT BE CH CY DE UK ES FI FR GB GR IE IT LU MC NL PT SE
W: AU JP US

AU 9946658 A 19981120 (200001) G01N027-327

ADT WO 9963333 A1 WO 1998-38421 19980531; AU 9946658 A AU 1998-46658 19990531

FDT AU 9946658 A Based on W. 9963333

PRAI US 1438-97402 19980519

IC ICM G01N027-327

IC S C10Q001-00; G01N033-543; G06F019-00

ICI G06F15/00

AB WO 9963333 A UPAB: 20000215

NOVELTY - A method for selecting an optimized regeneration solution for the regeneration of a biosensor surface having a surface-bound ligand and an analyte associated with the ligand, is new.

DETAILED DESCRIPTION - The method comprises:

(1) sequentially contacting the biosensor with first regeneration cocktails, where each one is an aqueous solution comprising at least one acidic, basic, ionic, organic, detergent or sanitizing stock solution, and where at least one of the cocktails comprises a mixture of at least two of the stock solutions;

(2) measuring the regeneration effect for each of the cocktails to determine which have the highest measured regeneration effect;

(3) selecting at least two different stock solutions present in the cocktails having the highest measured regeneration effect;

(4) combining the stock solutions in various ratios to generate another set of regeneration cocktails;

(5) sequentially contacting the biosensor surface with each of the new cocktails; and

(6) determining the regeneration effect of each cocktail, and identifying one as the optimized regeneration solution.

INDEPENDENT CLAIMS are also included for the following:

(1) a method for characterizing a ligand or analyte associated with a biosensor surface, comprising:

(a) contacting the biosensor surface, having a bound ligand, sequentially with characterization solutions;

(b) introducing the analyte into each solution, to interact with the bound ligand;

(c) measuring at least one of, association or dissociation rates, surface bound analyte concentration, and regeneration effect of the analyte and ligand interaction for each solution;

(d) characterizing the ligand or analyte from the results; and optionally

(e) comparing the characterization of the ligand and/or analyte associated with the biosensor surface with a set of predetermined characteristics of other test molecules, and predicting their activity; and

(2) a generated data signal, or computer memory containing a data structure, useful for communication of chemical perturbation information associated with an analyte-ligand interaction, comprising at least one kinetic parameter, each expressed as a mathematical model describing the relation between the parameter and analyte-ligand interaction in characterization solutions, so that the data can be used to communicate chemical perturbation information associated with the analyte-ligand interaction.

USE - Ligand-analyte binding pair interactions are used in assay work. They include antibody-antigen, hormone-hormone receptor, sense-antisense polynucleotides, avidin or streptavidin-biotin, enzyme-substrate or inhibitor, lectin-matching saccharide, lipid or polynucleotide -binding or matching protein, receptor-transmitter, drug target, protein-protein, DNA-DNA, and DNA-RNA. The process is used to break ligand-analyte interactions to regenerate a surface contaminated with analyte, for reuse in the next assay. Many surfaces suitable for carrying a required ligand are costly to produce, and must be reused if possible. As a notable example, the biosensor surface is a gold layer, which is capable of supporting surface plasmon resonance (SPR), or other optical detection and assay technique. The ligand is bound to this layer, preferably indirectly through a dextran matrix. The recovery can't be too harsh, or the ligand will be lost. The recovered analyte may be separated and collected for subsequent other assays if required. From a recovery profile from different solvents as determined with known samples, the analyte can also be characterized, even to detect minute differences, for quality control or other purposes. The interaction profile of ligand-analyte may be used to predict activity of the analyte, to be useful for screening. The regeneration and characterization procedures can be automated, using a computer. The computer is provided with a readable medium as a flowsheet, which instructs instrumentation to prepare the

various solutions (unless, initially, a kit is available), and following the conducting and assessment procedures as for manual control.

Dwg.0733

FS CPI EPI

FA AB

MC CPI: B11-C07A; B11-C08E; B11-C08E3; B11-C08E5; B11-C09; B12-K04F; D05-A02; D05-H09; D05-H10; D05-H11; D05-H12D; J04-B11
EPI: S03-E04A5S; S03-E04B5; S04-E14H4; S03-S14H9; T01-J03; T01-J06A1; T01-S01C; T01-S03

TECH UPTK: 200000019

TECHNOLOGY FOCUS - ORGANIC CHEMISTRY - Preferred components: The acidic solution is a mixture of acids having a pH < 7, and has a pH of 1-7. Specified acids include common inorganic, carboxylic and sulfonic acids, amino acids, phenols and purine acids, the most preferred acid is a mixture of oxalic, phosphoric, formic, and malonic acids. The basic solution is a mixture of bases having a pH > 14, examples include basic salts, ammonium, lead and zinc hydroxides, hydroxylamine and hydrazine, the most preferred is a mixture of ethan diamine, sodium phosphate, piperazine and glycine. The solvent is preferably a mixture of dimethyl sulfoxide, formamide, ethanol, acetonitrile and 1-butanol. The detergents anionic, cationic, zwitterionic and non-ionic, most preferably CHAPS, Swittergent 3-12 (RTM), Tween 80 (RTM), Tween 20 (RTM), and Triton X-100 (RTM). Chelating solutions include EDTA, EGTA, NTA, DOTA, GLEDTA, ETHEDTA, IDA and crown ethers, most preferably EGTA. A guiding concentration for each component is a minimum of 20mM. In the mixed solutions e.g. for RG2, nonpolar stock solutions may also be included. The characterization solutions are the same types (acidic, basic, ionic, organic, detergent and chelating) as the regeneration solutions and cocktails.

TECHNOLOGY FOCUS - INORGANIC CHEMISTRY - Preferred components: Ionic solutions are predominantly inorganic, the organic being acetic and haloacetics, and tetramethylammonium, most preferably potassium thiocyanate, magnesium chloride, urea and guanidinium chloride.

L123 ANSWER 26 OF 73 WPIX COPYRIGHT 2001 DEFVENT INFORMATION LTD
AN 2000-097578 [96] WPIX
DNN H2000-075173 DNC C2000-018203
TJ Novel **polypeptide** used to diagnose and treat bacterial infection.
DC B04 D16 S03 T01
IN BLACK, M T; JAWORSKI, D D; KOSMATKA, A L; SHILLING, L K; WANG, M; WILDING, E I
PA (SMIK) SMITHKLINE BEECHAM CORP
CYC 19
PI WO 9962527 A1 19991209 (200006)* EN 57p A61K035-74
RW: AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE
W: JP
ADT WO 9962527 A1 WO 1999-US12301 19990606
PRAI US 1998-92218 19980605
IC ICM A61K035-74
ICS A61K038-04; C07K016-00; C12H001-00; C12N015-00; C12N015-31; C12P021-02; C12Q001-64; G01N073-68; G06F017-30; G06F159-00
AB WO 9962527 A UPAB: 200000019
NOVELTY - An isolated **polypeptide** (A) selected from: a **polypeptide** comprising at least 70, 80, 90 or 95% identity to the 719 residue amino acid sequence fully defined in the specification, a **polypeptide** which comprises, or is, the 719 residue sequence, and a **polypeptide** encoded by a recombinant polynucleotide comprising the 2160 bp DNA sequence fully defined in the specification, is new.
DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:
(1) an isolated polynucleotide selected from: a polynucleotide comprising a polynucleotide sequence encoding (A), a polynucleotide comprising a sequence that has at least 70, 80, 90 or 95% identity to a

polynucleotide sequence encoding the 719 residue sequence, a polynucleotide sequence comprising at least 70, 80, 90 or 95% identity to the 2160 bp DNA sequence, a polynucleotide comprising a sequence encoding the 719 residue sequence, a polynucleotide which is the 2160 bp DNA sequence, an isolated polynucleotide obtained by screening a library under stringent hybridization conditions with a probe having the 160 bp DNA sequence or a fragment of it, a polynucleotide encoding a mature **polypeptide** expressed by the *nrdE* gene contained in *Streptococcus pneumoniae*, and a polynucleotide sequence complementary to any of the isolated polynucleotides above;

(i) an antibody antigenic or immuno-specific for (A);

(ii) a method for the treatment of an individual:

(a) in need of enhanced activity or expression of (A) comprising administering to the individual an agonist to the **peptide**; or providing to the individual an isolated polynucleotide comprising the polynucleotide sequence encoding the **polypeptide** in a form so as to effect production of the **polypeptide** activity *in vivo*; or

(b) having need to inhibit the activity or expression of (A), comprising, administering to the individual an antagonist to the **peptide**, a nucleic acid molecule that inhibits the expression of a polynucleotide sequence encoding the **polypeptide**, or a **polypeptide** that competes with (A) for its ligand, substrate or receptor;

(4) a process for diagnosing or prognosis a disease or susceptibility to a disease in an individual related to expression or activity of (A) comprising determining the presence or absence of a mutation in the nucleotide sequence encoding (A) in the **genome** of the individual, or analysing for the presence or amount of (A) expression in a sample derived from the individual;

(5) a method for screening to identify compounds that activate or inhibit the **function** of (A) comprising:

(a) measuring the binding of a candidate compound to (A) or to cell membranes bearing it or a fusion **protein** thereof by means of a label directly or indirectly associated with the candidate compound, or in the presence of a labeled competitor;

(b) testing whether the candidate compound results in a signal generated by activation or inhibition of (A), using detection systems appropriate to the cells or cell membranes bearing (A);

(c) mixing a candidate compound with a solution containing (A), to form a mixture, measuring activity of (A) and comparing the activity of the mixture to a standard;

(d) detecting the effect of a candidate compound on the production of mRNA encoding (A) and (A) in cells, using for instance an ELISA assay; or

(e) contacting a composition comprising (A) with the compound to be screened under conditions to permit interaction between the compound and (A) to assess the interaction of a compound, such interaction being associated with a second component capable of providing a detectable signal in response to the interaction of (A) with the compound, and determining whether the compound interacts with and activates or inhibits an activity of (A) by detecting the presence or absence of a signal generated from the interaction of the compound with (A);

(6) an agonist or antagonist of the activity or expression of (A);

(7) an expression system comprising a polynucleotide capable of producing (A) when the expression system is present in a compatible host cell;

(8) a host cell comprising the expression system of (7) or a membrane of it expressing (A);

(9) a process for producing (A) comprising the step of culturing the host cell of (8) under conditions sufficient for the production of (A);

(10) a process for producing a host cell comprising the expression system of (7) or a membrane of it expressing (A) comprising the step of transforming or transfecting a cell with an expression system comprising a polynucleotide capable of producing (A) when the expression system is present in a compatible host cell, under appropriate conditions, produces (A);

(11) a host cell produced by the process of (10) or a membrane

thereof expressing (A);

(13) a computer readable medium having stored thereon: a polynucleotide comprising the 2160 bp DNA sequence, a **polypeptide** comprising the 719 residue sequence, a set of polynucleotide sequence where at least one of the sequences comprises the 2160 bp DNA sequence, a set of **polypeptide** sequences where at least one of the sequences comprises the 719 residue sequence, a data set representing a polynucleotide sequence comprising the 2160 bp DNA sequence, and encoding a **polypeptide** sequence comprising the 719 residue sequence;

(13) a computer based method for performing homology identification, comprising providing a polynucleotide sequence comprising the 2160 bp DNA sequence in a computer readable medium and comparing the polynucleotide sequence to at least one polynucleotide or **polypeptide** sequence to identify homology; and

(14) a computer based method for polynucleotide assembly, the method comprising providing a first polynucleotide sequence comprising the 2160 bp DNA sequence in a computer readable medium, and screening for at least one overlapping region between the first polynucleotide sequence and a second polynucleotide sequence.

ACTIVITY - Antibiotic; antiinflammatory; auditory; antialcer.

MECHANISM OF ACTION - Vaccine.

USE - The rrdE sequences may be used in a diagnostic method for diagnosis of disease, staging of disease or response of an infectious organism to drugs. The sequences are preferably used to diagnose bacterial infections, particularly those caused by *Streptococcus pneumoniae*. The sequences may also be used to prevent *Helicobacter pylori* infection, and stomach cancer, ulcers and gastritis caused by *H. pylori* infection. They may also be used to treat diseases caused or related to infection by bacteria, including otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema, and endocarditis.

AVAILABILITY - None given.

Dwg.0/0

FS CPI EPI
 FA AF; DCN
 MC CPI: B04-C01G; B04-E02F; B04-E05; B04-E06; B04-E08; B04-F010E;
 B04-N03A0E; B11-A01; B11-C03E5; **B12-K04A4; B12-K04F**
 ; B14-A01B2; B14-F01; B14-J01; B14-K01; B14-N03; B14-N04; D05-H08;
D05-H09; D05-H11; D05-H12A; D05-H13D1; D05-H13D2; D05-H13E;
 D05-H13; D05-H14; D05-H17A6
 EPI: S03-E14H4; S03-E14H5; T01-J10C4

L123 ANSWER 29 OF 73 WPIX COPYRIGHT 2001 FERWENT INFORMATION LTD

AN 2000-038624 [03] WPIX

DNN N2000-029163 DMC C2000-009847

TI Three dimensional computer modeling to identify HLA binding compounds useful for modulating immune responses.

DC B02 B05 D16 **T01**

IN KOHLEF, N; LIU, M; RICHERT, J R; WANG, S; WU, X; YIN, D
 PA (GEOU) UNIV GEORGETOWN

CYC 35

PI WO 9955681 A1 19991104 (200003)* EN 64p C07D239-72
 EW: AT BE CH CY DE DK EA ES FI FP GB GR GM GE IE IT KE LG LU MC MW NL
 OA PT SD SE SL SE UG SW
 W: AE AL AM AT AU AZ BA BB BG BE BY CA CH CN CU CS DE DK EE ES FI GB
 GD GE GH GM HE HU ID IL IN IS JP KE KG KP KR KE LC LY LR LS LT LU
 LV MD MG MK MN MW MK NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR
 TT UA UG US VN YI SA SW

AU 9936691 A 19991116 (200015) C07D239-72

ADT WO 9955682 A1 WO 1999-US9218 19990429; AU 9936691 A AU 1999-36691 19990429

FIGT AU 9936691 A Based on WO 9955681

PRAI US 1998-83426 19980419

IC ICM C07D239-72

ICG C07D215-16; C07D217-00; C07D217-10; C07D217-22; **G06F019-00**

AB WO 9955682 A UPAB: 20000118

NOVELTY - Methods (A) for identifying and using HLA (Histocompatibility Lymphocyte-A System) binding compounds as HLA-agonists and antagonists,

comprising the computational processing of a database containing 3-dimensional structures of receptor sites and chemical compounds, are new.

DETAILED DESCRIPTION - Method (A) of processing a compound data base containing 3-dimensional structures of chemical compounds to provide a lead compound capable of blocking a receptor site in a host molecule comprising:

(1) modeling the 3-dimensional structure of the receptor site;

(2) positioning a compound from the data base in the receptor site and assigning a geometrical-fit score, indicating the fit between the compound and the receptor site;

(3) ranking the compound in the data base according to their score, and forming a group of compounds with a rank of a predetermined value, or higher;

(4) minimizing an energy function describing interactions between a compound and a receptor site by adjusting coordinates of the compound to obtain a minimum-energy compound-host molecule complex structure;

(5) ranking the compounds according to their minimum energy values and forming a subgroup of compounds with a minimum-energy rank of a predetermined value or higher; and

(6) visualizing (on a computer) a minimum energy compound-host molecule complex and forming a second subgroup of compounds with a visual-fit satisfying a predetermined criterion.

An INDEPENDENT CLAIM is also included for a method of inhibiting the interaction of an HLA molecule to an antigen comprising the administration of at least one compound of formula (I).

P1, P2 = optionally substituted phenyl, benzyl or other 5- or 6-membered aromatic ring system, optionally containing one or more heteroatoms selected from O, S and N

P3, P4 = H, optionally substituted phenyl, benzyl or other aromatic ring system, 1-10C alkyl, 1-10C alkoxy, halogen, SO2M, amide, or COOR

M = H or alkyl

P5, P6, R7, R8 = H, halogen (F, Cl, Br, I), alkyl, 1-10C alkoxy, amide, NO2, amine, 1-10C cycloalkyl, nitroso, OH, ether, ester, sulfonic acid, alkenyl or allyl

X, Y = N or C

ACTIVITY - Immunosuppressive.

MECHANISM OF ACTION - HLA-agonists and antagonists.

USE - Compounds identified by (A), or of formula (I) can be used to treat autoimmune diseases, graft versus host disease, transplant rejection and **multiple** sclerosis.

ADVANTAGE - (A) allows adjusting of a compound's structure to optimize the fit between the host molecule and homologous compound. The method also allows for the modeling of host **proteins** whose 3-D structure is unknown.

DESCRIPTION OF DRAWING(S) - The diagram shows a 3-D model of the HLA-DR₀₁ molecule produced by homology modeling.

Dwg.1/10

FS CPI EPI

FA AE; GI; DSN

MC CPI: B04-B04C2; B04-C01; B04-E01; B04-G04; B04-G05; B04-N01; B06-D04; B06-D07; B10-A03C; B11-C01; B14-G02; B14-G03C; B14-G03D; B14-L01; B14-L06; B14-L01; D05-H09; D05-H11

SPI: T01-J05B4B; T01-J05R4P; T01-J06A; T01-J10C4

TECH UPTX: 20000118

TECHNOLOGY FOCUS - PHARMACEUTICALS - Preferred Treatment: Compounds of formula (I) are further administered with another active agent selected from CS40-ligand antagonist, soluble CS40, anti-cytokine antibody or anti-cytokine receptor antibody. HLA-inhibition results in reduced cytokine, preferably IL-1, production.

Preferred Compound: (I) is 2-((4-(acetylamino)phenyl)amino)-N-((4-(acetylamino)phenyl)amino)sulfonyl)-4-oxo(3-hydroquinazolin-3-yl)acetamide or one of 15 disclosed analogues, especially 2-((2,4-dichlorophenyl)amino)-N-(7-((2,4-dichlorophenyl)amino)sulfonyl)-1-

exo(hydronaphthyl)acetamide.

TECHNOLOGY FOCUS - BI-TECHNOLOGY - Preferred Energy Function:

The energy **function** of (4) comprises a Van der Waals interaction term and an electrostatic interaction term. Minimizing the energy **function** comprises probing the compound's conformational flexibility. The minimum energy is a global minimum of the **function**. Modeling the 3-dimensional structure of the receptor sites comprises providing a set of three coordinates for each atom of the host molecule (preferably HLA-DR1301) defining a position of the center of the atom in a 3-dimensional referential.

Preferred Processing Method: The host molecule is a **protein** with known primary and secondary structures, and unknown tertiary structure. The position of the center of each atom of the host molecule is provided by:

(a) aligning the host molecule sequence with a homologous **protein** (preferably DR1 with a tertiary structure defined by an X-ray structure of DR1 complexed with an influenza **peptide**) sequence obtained from a database of **proteins** with known tertiary structures;

(b) assigning a sequence-homology score to each homologous **protein**, indicating the percentage of amino acids occupying identical positions in the host and homologous sequences, and forming a template host tertiary structure by overlaying the atoms of a host molecule backbone on the backbone of a homologous **protein** with a sequence-homology score of a predetermined value or higher, and overlaying the atoms of a host side chain with an equivalent side chain in the homologous **protein** on the corresponding atoms in the homologous **protein**; and (c) refining the tertiary structure template by adjusting the positions of atoms in a host side chain having an equivalent side chain in the homologous **protein** to provide a structure with a low energy value defined by interactions between the atoms in the host molecule. Refining the template tertiary structure preferably comprises positioning a template compound with known host-binding properties in the receptor site, and adding a term describing interactions between the template compound and a host molecule side chain to the internal energy **function**. The method further comprises testing identified compounds for their ability to specifically bind to HLA-DR1301 and selecting those with the greatest binding affinity. Analogs of the selected compounds are obtained, and are tested by an in vitro assay that measures binding to HLA-DR1301, and those with the greatest affinity are selected for in vivo usage.

Preferred Receptor Site: The receptor site comprises a negatively charged pocket and a hydrophobic pocket, and the template compound comprises the side chains of residues 114 and 162 of Myelin Basic **Protein**. The side chain of residue 114 is disposed in the hydrophobic pocket and the side chain of residue 162 is in the negatively charged pocket.

L123 ANSWER 30 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 2001-038354 (03) WPIX

DNN US 600-028949 DMC C 000-0009744

TI Novel Staphylococcal RNAI polynucleotides and **polypeptides** useful for screening for antibacterial compounds.

DC B04 D16 D21 D22 S03 T01

IN BLACK, M T

PA JAZMIK SMITHKLINE BEECHAM CORP

CYC 24

PI WO 910147 A1 1999-09-13 (00001) EN 55p A61K035-74

RW: AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE

W: JP

EP 1064007 A1 2001-0103 (200101) EN A61K035-74

R: BE CH DE DK FR GB IT I I NL

ACT WO 947147 A1 WD 1993-03-15; EP 1064007 A1 EP 1999-912551 1993-03-15, WO 1999-03-15; EP 1064007 A1 EP 1999-912551 1993-03-15, WO 1999-03-15

FLT EP 1064007 A1 Based on WO 947147

PRAI US 1998-40213 19980317

IC ICM A61F035-74

IC3 A61F038-04; C07K016-00; C12N001-00; C12N016-00; C12N015-31;
C12P021-02; C12Q001-68; G01N653-68; G06F017-30;**G06F159-00**

AB WO 9347147 A UPAB: 20000118

NOVELTY - *nrdG* polynucleotide (II) isolated from *Staphylococcus aureus* and the **polypeptide** (I) it encodes, is new.DETAILED DESCRIPTION - The isolated *nrdG* polynucleotide (II) selected from:

(a) an isolated polynucleotide that has at least 95% identity to, and over the entire length of, a nucleotide sequence encoding the 173 amino acid sequence (S1) given in the specification;

(b) an isolated polynucleotide comprising a nucleotide sequence which has at least 95% identity to the 514 basepair sequence (S2) given in the specification;

(c) an isolated polynucleotide that has or is the S2 sequence;

(d) an isolated polynucleotide obtainable by screening an appropriate library under stringent hybridization conditions with a probe having the S2 sequence, or its fragment;

(e) an isolated polynucleotide encoding a mature **polypeptide** expressed by the *nrdG* gene of *S. aureus*; and

(f) a polynucleotide sequence complementary to the polynucleotides of

(a) to (c).

INDEPENDENT CLAIMS are also included for the following:

(1) An isolated **polypeptide** selected from:(a) an isolated **polypeptide** comprising an amino acid sequence having at least 35% identity to S1;(b) an isolated **polypeptide** comprising S1;(c) an isolated **polypeptide** which is S1;(d) An antibody antigenic to or immunospecific for the **polypeptide** of (1);

(2) A method for the treatment of an individual in need:

(a) of enhanced activity or expression of (1);

(b) to inhibit activity or expression of the **polypeptide** of (1);

(3) A process for diagnosing or prognosis a disease or a susceptibility to a disease in an individual related to expression or activity of (1), comprising:

(a) determining the presence or absence of a mutation in (II) in the **genome** of the individual; and/or

(b) analyzing for the presence or amount of expression of (II) in a sample derived from the individual;

(5) A method for screening to identify compounds that activate or that inhibit the **function** of the **polypeptide** of (1);(6) An agonist or an antagonist of the activity or expression of the **polypeptide** of (1);(7) An expression system comprising a polynucleotide capable of producing the **polypeptide** of (1) when the expression system is present in a compatible host cell;(8) A host cell comprising the expression system of (7) or its membrane expressing the **polypeptide** of (1);(9) A process for producing the **polypeptide** of (1) comprising culturing the host cell of (8) under conditions sufficient for the production of the **polypeptide**;(10) A process for producing a host cell comprising the expression system of (7), comprising transforming or transfecting a cell with an expression system such that the host cell, under appropriate culture conditions, produces the **polypeptide** of (1);(11) A host cell produced by the process of (10) or its membrane expressing **polypeptide** of (1);

(12) A computer readable medium having stored on it a member selected from the group consisting of:

(a) a polynucleotide comprising the S2 sequence;

(b) a polypeptide comprising the S1 sequence;

(c) a set of polynucleotide sequences wherein at least one of the sequences is S2;

- (d) a set of polypeptide sequences wherein at least one of the sequences comprises the S1 sequence;
- (e) a data set representing a polynucleotide sequence comprising the S2 sequence;
- (f) a data set representing a polynucleotide sequence encoding a polypeptide sequence comprising the S1 sequence;
- (g) a polynucleotide comprising the S2 sequence;
- (h) a polypeptide comprising the S1 sequence;
- (i) a set of polynucleotide sequences where at least one of the sequences comprises the S1 sequence;
- (j) a set of polypeptide sequences where at least one of the sequences comprises the S1 sequence;
- (k) a data set representing a polynucleotide sequence comprising the S2 sequence;
- (l) a data set representing a polynucleotide sequence encoding a polypeptide sequence comprising the S1 sequence; and

(13) A computer based method for performing homology identification, comprising providing a polynucleotide sequence comprising the S2 sequence given in the specification in a computer readable medium, and comparing the polynucleotide sequence to at least one polynucleotide or polypeptide sequence to identify homology.

ACTIVITY - Antibacterial; Cytostatic; Veterinary.

MECHANISM OF ACTION - Vaccine.

USE - The polynucleotides and polypeptides may be employed as research reagents and material for the discovery of treatments and diagnostics for diseases, particularly human diseases. They can be used for diagnosis of the disease and staging of disease, and as reagents in differential screening methods. The polynucleotides may be used as a source for hybridization probes, and for screening of genetic mutations, serotype, and identification, and for organism chromosome identification. The polypeptides can be used to produce antibodies. The polypeptides can also be used in vaccine formulations, and to identify agonists and antagonists. These are used to prevent, inhibit or treat diseases, particularly of *Helicobacter pylori* infections, such as gastrointestinal carcinoma. Diseases diagnosed, treated and prevented by the products include infections of the upper and lower respiratory tract, cardiac, gastrointestinal, CNS, eye, kidney and urinary tract, skin, bone and joint. The polypeptides can also be used to treat wounds and body implants to prevent infection.

ADVANTAGE - The frequency of Staphylococcal infections has risen dramatically, and it is no longer common to isolated *S. aureus* strain that are resistant to standard antibiotics. The nrdG products of the invention can be used screen for new antibacterial compounds that may target these resistant bacteria.

Dwg.0/0

FS CFI EPI
 FA AE; DGN
 MC CFI: B04-C01G; B04-E03F; B04-E05; B04-E08; B04-F10A3E; B04-F1100E;
 B04-G01; B04-N03A; **B12-K04A**; B14-A01; B14-H01; B14-J01;
 B14-L01; B14-L06; B14-N01; B14-N03; B14-N07; B14-N10; B14-N17B;
 B14-S11B; D05-A01A4; D05-A01B; D05-H07; D05-H08; **D05-H09**;
 D05-H11; D05-H11A; D05-H12D1; D05-H12E; D05-H14; D05-H17A6; D08-B09;
 D09-C01D
 EPI: S03-E14H; T01-J05B

L123 ANSWER 31 OF 73 WPIX COPYRIGHT 2001 SERVENT INFORMATION LTD

AN 1999-579903 (49) WPIX

DNN 01399-428125 DDC C1999-103667

TI Identifying pathways of drug action.

DC B04 B16 S03 **T01**

IN FPIEND, S H; STOUGHTON, R

PA (ROCHE-N) ROSETTA INPHARMATICS INC

CYC S6

PI US 5965352 A 19991012 (199949)* 47; C12001-00

WO 9953708 A1 19991116 (200001) EN C12001-00

FW: AT BE CH CY DE DK EA ES FI FF GB GH GM GR IE IT KE LS LU MC MW NL

OA PT SD SE SL SE UG SW

W: AE AL AM AT AU AS BA BB BG BR BY CA CH CN CU C.U DE DK EE ES FI GB
GD GE GH GM HF HJ IP IL IN IS JP KE PG KP KR KU LC LF LS LT LU
LV MD MG MK MN MW MX NO NS PL PT EO PU SD SE SG SI SK SL T.U TM TR
TT UA UG US VN YU ZA SW

AU 9928906 A 19991119 (200014) C12Q001-00

ADT US 5965382 A US 1998-74983 1998-505; WO 9958707 A1 WO 1999-US1005-19980507; AU 9938906 A AU 1999-74986 1999-507

FDT AU 9938906 A Based on WO 9958707

FRA1 US 1998-74983 19980508

IC ICM C12Q001-00

ICP AG1F045-00; C11N011-00; C11Q011-00; C07D19-00; G06F017-00;

G06F017-14

AB US 5965382 A UPAK: 19991124

NOVELTY - Methods for determining the primary and secondary biological pathways through which drugs act on cells and identifying the **proteins** and genes which are affected via each pathway, are new. Methods are also included for identifying the pathways affected by environmental change, disease and those through which the side-effects of drugs are mediated.

DETAILED DESCRIPTION - A method (I) for identifying the biological pathways involved in the action of a drug on a cell, comprises:

(i) determining the cellular response produced in the cell by different levels of exposure to the drug (by measuring its effects on a number of cellular constituents);

(ii) representing a model of drug response as a combination of biological pathway responses (produced by measuring the cellular constituents of a biological pathway in response to different levels of perturbation) in the cell, one of which is subject to an independent scaling transformation; and

(iii) determining best scaling transformations of the biological pathway responses which minimize the value of an objective **function** of the difference between the candidate drug response and the model drug response (the combination of the responses subject to the best scaling transformation identifies the biological pathways involved in the action of the drug on the cell type).

Other INDEPENDENT CLAIMS are included for similar methods for:

(1) determining a more pathway-specific drug candidate from an initial drug candidate;

(2) identifying specific biological pathways which are involved in the action of a drug and which mediate side-effects of the drug;

(3) identifying a number of specific biological pathways that are involved in mediating therapeutic efficacy for a disease or disorder;

(4) measuring the similarity of the effects of two drugs on a cell type;

(5) identifying biological pathways involved in the effects of an environmental change upon a cell type; and

(6) identifying biological pathways involved in the effects of an environmental change upon a cell type.

USE - The methods may be used to characterize the effects of drugs on cells, in particular for finding biological pathways in cells affected by drug action and by environmental change and disease. This information is useful in the discovery of new drugs and alternative therapies and in understanding efficacy, side effects, toxicities and activation of metabolic responses.

Dwg.073

FS CPI EPI

FA AB; DCN

MC CPI: B04-F01; B11-A; B11-C03E1; B11-C03E2; B12-K04A;
B12-K04E; D05-A04; D05-H03; D05-H09

EPI: S03-E14H4; T01-J04B1

TECH UPTK: 19991124

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preferred Method: In (I), step (iii) comprises determining an actual minimized value of the objective **function**. This is determined from the candidate drug response and the model drug response. Therefore, (I) further comprises a step of

assessing the statistical significance of the best scaling transformations in a number of pathways. This comprises:

(i) obtaining an expected probability distribution of the minimized values of objective **function**; and

(ii) assessing the statistical significance of the actual minimized value of objective **function** in view of the expected probability distribution of minimum values of objective **function**.

The expected probability distribution of the minimum values of objective **function** comprises:

(i) randomizing the drug responses at different levels of drug exposure and randomizing the model drug response by randomizing a number of biological pathway responses at different levels of perturbation;

(ii) determining a theoretical minimum value of the objective **function** by finding best scaling transformations of a number of the randomized pathway responses which minimize the objective **function** of the difference between the randomized responses and the randomized model responses; and

(iii) repeating steps (i) and (ii) to determine a number of theoretical minimum values (which form an expected probability distribution of minimized values).

The method (I) may further comprise verifying that the biological pathways are actually involved in the action of the drug by selecting a model response (either the first or second model response) which behaves most similarly to a combined drug perturbation response. The drug perturbation response is provided by measuring a number of cellular constituents in a cell exposed simultaneously to varying levels of drug exposure and varying levels of perturbations within the biological pathways. The first and second model responses comprise a combination of several pathway responses subject to best scaling transformations evaluated at a number of sums (each sum is the sum of a number of levels of drug exposure subject to the scaling transformations and varying levels of perturbations to the pathways). The biological pathways are verified as pathways involved in the action of the drug if the first model response is used.

(I) may also comprise assigning a cellular constituent present in the drug response to one or more of the biological pathways (the biological pathway response of the constituent subject to its best scaling transformation has the greatest correlation with the drug response of the cellular constituent). The scaling transformations comprise transforming levels of drug exposure to corresponding levels of perturbation to the biological pathways. Transforming is done by linear mapping. The objective **function** is minimized by selecting best sets of parameters for the scaling transformations. The objective **function** comprises a sum of the squares of differences in drug responses at varying levels of exposure to the drug and the model response at differing levels of exposure (further subject to scaling transformations). The objective **function** comprises the negative of the correlation of the drug response and the model drug response.

The biological pathway responses are produced by a method comprising interpolating the measured values (by approximating a sum of spline **functions** or a Hill **function**). The pathways tested are

chosen from a compendium of those likely to be involved in drug responses.

The cellular constituent measured is preferably RNA, measured by contacting a gene transcript with RNA of the cell or with cDNA derived from it. The transcript array comprises a surface with attached nucleic acids or mimics which are capable of hybridizing to the RNA (or cDNA). The RNA levels in drug-treated and untreated cells is compared.

Protein levels may also be measured using antibody arrays (comprising a surface with attached antibodies which bind to the **proteins**) or by 2 dimensional electrophoresis.

The activities of the cellular constituents may also be measured. The biological pathways originate at a number of specific cellular constituents under the control of an expression system. Perturbations are caused by altering the expression and activity constituents (e.g. by transfection with a heterologous gene or by the action of ribozymes and antisense sequences). The expression of the constituents may also be altered by the expression of the **protein** as a fusion

protein comprising the **protein** and a degron (which degrades the **protein**), or by the use of antibodies to the **protein** or other antagonists, drugs and dominant negative mutant **proteins**.

Preferred Cells: The cells are substantially isogenic to *Saccharomyces cerevisiae*.

L123 ANSWER 32 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 1999-570766 [46] WPIX

CNN N1999-420477 EPO C1999-166577

TI Predicting the folded structure of **proteins**.

DC B04 D16 N04 S03 T01

IN BENNEF, J A

PA (BENN-I) BENNEF J A

CYC 1

PI US 5955764 A 19990816 (1+9945)* 113p G01N33-00

ADT US 5955764 A US 1992-657224 19920325

PRA1 US 1992-657224 19920325

IC ICM G01N033-00

ICS G06F015-00

AB US 5955764 A UPA8: 19991122

NOVELTY - Predicting the folded structure of **proteins**, by aligning sequences of homologous **proteins** and using patterns of **evolutionarily** conserved and varied sequences to assign positions, is new.

DETAILED DESCRIPTION - This method is used for predicting the folded structure of **proteins**, comprising aligning the sequences of homologous **proteins**, using patterns of conservation and sequence variation with clearly defined **evolutionary** relationships. Positions in the alignment are assigned to the surface or inside of the folded structure, active sites, and parsing segments. Secondary structural units are assigned by identifying periodicity in the assignments, and assembled into globular form using distance constraints imposed by disulfide bridges, active site assignments and co-variation analysis.

An INDEPENDENT CLAIM is also included for a method predicting the secondary structure of **proteins** using the above method up to the point of assigning secondary structural units.

USE - The predicted secondary structures are useful for identifying antigenic sites on a **protein** molecule, as guides for site directed mutagenesis studies, and for understanding the interaction of a **protein** with other molecules.

ADVANTAGE - The method is more efficient than prior art methods of predicting folded structure of **proteins**.

Dwg.0728

FS CPI EPI

FA AB; DCN

MC CPI: B04-N04; B11-C08E; B12-M04E; C05-H03; G04-B01

EPI: S03-E14; T01-7

TECH UPTX: 19991122

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Essential features: Essential features of the method are:

- (i) examining aligned sequences of several homologous **proteins** rather than a single sequence of a single **protein**;
- (ii) extracting information concerning the three dimensional structure of the **protein** family from patterns of conservation and variation within a set of homologous sequences, rather than by a simple averaging of a property of the sequences taken individually;
- (iii) combining algorithms that assign positions in the alignment to the surface or interior of the folded structure, and to the active site, as a first step for predicting secondary structural elements;
- (iv) identifying separate secondary structural elements in the alignment, using parsing algorithms that identify gaps in the alignment and specific parsing sequence motifs;
- (v) the algorithms are applied to subgroups of **proteins** with clearly identified **evolutionary** relationships, in particular a clearly specified sequence identity and **evolutionary** distance;

(vi) the algorithms are designed to reflect how natural selection and neutral drift influence the divergent **evolution** of **protein** sequences; and

(vii) assembling the secondary structural elements to form super secondary and tertiary structural models by orienting these elements using disulfide bridges, active site assignments and co-variation analysis.

L123 ANSWER 33 OF 73 WPIX COPYRIGHT 2001 DEFWENT INFORMATION LTD

AN 1999-462430 [39] WPIX

DNN N1999-346199 DDC C1999-136174

TI Method for designing artificial **protein** - useful for producing **proteins** with required functions.

DC B04 D16 T01

PA (FIKA) FIFAGAKU KENKYUSHO

CYC 1

PI JP 11193397 A 1994-07-01 (1999) * 10p C07K014-00

ADT JP 11193397 A JP 1993-128752 19951006

PRAI JP 1997-272431 19971006

IC ICM C07K014-00

ICS C07B061-00; C07K014-805; C12N015-09; G06F017-30

AB JP 11193397 A UPAB: 19991004

A method for producing an ideal amino acid sequence comprises the following steps: (1) preparation of one initial amino acid sequence corresponding to the structure of a **protein**, (2) selecting the amino acid residues optimum to each site; (3) repeating the steps of selecting the amino acid residues optimum to each sites of the total amino acid residues constituting the N-order amino acid sequence (N is an integer not less than 1) to give an N+1-order amino acid sequence consisting of the selected amino acid residues until the N-order amino acid sequence comes to be same as the N+1-order amino acid sequence, and (4) selecting the resultant N-order amino acid sequence as the optimum amino acid sequence of said **protein**.

Also claimed is a **protein** having an amino acid sequence obtained by the above method.

ADVANTAGE - The method can design a **protein** with desired functions and properties.

Dwg.0/5

ES CPI EPI

FA AB

MC CPI: B04-C01; B11-C08; D05-H09

EPI: T01-J05B; T01-J16C4

L123 ANSWER 34 OF 73 WPIX COPYRIGHT 2001 DEFWENT INFORMATION LTD

AN 1999-458333 [38] WPIX

DNN N1999-342838 DDC C1999-134564

TI Use of independent assay controls to perform and monitor assays, particularly immunoassays.

DC B04 D16 J04 S03 T01

IN ANDERPEFG, J M; BUECHLER, K F; MCPHERSON, F H

PA (BIOS-N) BIOSITE DIAGNOSTICS INC

CYC 85

PI WO 9935602 A1 1999-07-15 (1999) * EN 148p G06F019-00

EW: AT BE CH CY DE DK EA ES FI FF GB GH GM GF IE IT FE LS LU MC MW NL
OA PT SD SE SJ SG ZW

W: AL AM AT AU AZ BA BE PG PF BY CA CH CN CU CL DE DE EE ES FI GB GD
GE GH GM HF HU ID IL IN IS JP KE PG FF FG LC LE LS LT LU LV
MD MG MK MN MW MY NO NC PL PT RO FU SD SE SG SI SF SL TJ TM TR TT
UA US US VI YU ZW

AG 9921066 A1 1999-07-06 (1999) * 1393521 G06F019-00
EP 1046122 A1 2000-10-07 (2000) * EN G06F019-00

R: AT BE CH CY IE DK ES FI FF GB GR IE IT LI LU MC NL PT SE

ADT WO 9935602 A1 WO 1999-08-06 1999-0104; AU 9921066 A AU 1999-21066 19990104;
EP 1046122 A1 EP 1999-09-1345 1999-0104, WO 1999-US261 19990104

FDT AU 9921066 A Based on WO 9935602; EP 1046122 A1 Based on WO 9935602

PRAI US 1998-0965 19980105

IC ICM G06F019-00

AB ICS G01N033-53
 WO 9925602 A UPAB: 19990922
 NOVELTY - Use of independent assay controls (IACs) for monitoring and performing assays.

DETAILED DESCRIPTION - Method for determining the rate of flow of solution through an assay device comprising a reaction chamber (RC) and at least one diagnostic line (DL) comprises

- (i) providing a labeled member of a specific binding pair (MBP) in RC and a second MBP, bound to a solid phase, in DL, where neither MBP binds to assay reagents in the device, and
- (ii) determining the rate of flow from the level of signal, from the label, detected in DL.

INDEPENDENT CLAIMS are also included for the following:

- (a) similar methods for determining, in an assay device, the environmental conditions, progress and time to completion of the assay, deviant assay results, also for smoothing results and for verifying location of a detection zone;
- (b) apparatus and kits for these processes;
- (c) methods for determining a corrected assay result from a measured result and one or more
- (d) IAC results; and
- (e) computer programs for performing calculations involved in (c).

ACTIVITY - None given.

MECHANISM OF ACTION - None given.

USE - IAC are used to monitor and perform immunoassays, particularly to correct results; detect deviant results and to determine rate of flow, environmental conditions, progress and time to completion of the assay, also for smoothing an assay signal and to verify location of detection zones.

ADVANTAGE - The use of IAC ensures that assay results are correct, even when the assay conditions vary. IAC are independent of the presence/concentration of assay reagents, but depend on the sample matrix and progress of the immunoassay reaction, so can verify that this reaction has occurred without error.

Dwg.0/18

FS CPI EPI

FA AB; DCM

MC CPI: B04-C01C; B04-G01; B11-C07A; B12-K04; B12-K04E;
 D05-H09; J04-B01
 EPI: S03-E14H4; T01-J

TECH UPTK: 19990922

TECHNOLOGY FOCUS - INSTRUMENTATION AND TESTING - Preferred materials: MBP, and the assay reagents, are binding **proteins**, antibodies (or their fragments), **peptides** or organic molecules, and labels are dyes, colloidal solutions, enzymes or molecules that generate an electrical and/or magnetic signal. Preferred processes: To determine environmental conditions, both MBP (one labeled, the other with an affinity tag) are combined with the test sample and DL includes an affinity tag partner (atp). Environmental conditions are then related to the signal detected in DL. Particularly MBP are attached to the iid and/or the base of RC and atp is immobilized in DL. To measure progress/time to completion, a label, preferably linked to MBP, is present in RC and a signal from the label is detected in a selected region of DL. Analysis is based on the absolute value of this signal or its rate of change. To detect a deviant assay result, a similar system is used, and an assay result (AS) and an IAC signal are detected, in at least two separate zones of DL, and their shapes compared. To determine a corrected result (Tc), from a measured result (Tm) and j measurements of an IAC, the difference between each IAC and a mean value of IAC is determined and the difference, for each value, multiplied by a constant (β eta) and the products added together. The total is subtracted from Tm, or a quotient between Tm and this total is determined. Each β eta is determined by a matrix function or by linear regression from a plot of differences between assay results and mean value of these results vs. differences between IAC and mean value for IAC. Alternatively, (a) a function f , is determined for each IAC and a quotient calculated between Tm and

1+f, or (b) a difference between each IAC and a mean value of IAC is determined, **multiplied** by a constant, the sum of 1+ every j product calculated, and a quotient between Tm and this sum determined. The j IAC values are particularly signals measured in the new processes.

L123 ANSWER 25 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 1999-458146 [38] WPIX
 DNN N1999-342695 INC C1999-134481
 TI Predicting presence of abnormal levels of **protein** in blood clotting cascade.
 DC B64 SOC **T01**
 IN BFAUH, P; FISCHER, T J; GIVENS, T B
 PA (ALKU) AFSC NOBEL NV
 CYC 24
 PI WO 9934208 A1 19990708 (199908) * EN 94p G01N 23-49
 PW: AT BE CH CY DE DK ES FI FF GB GR IE IT LU MC NL PT SE
 W: AU CA JP KR US
 AU 9919503 A 19990719 (199951) G01N 23-49
 EP 1042669 A1 20001011 (200052) EN G01N 23-49
 E: AT BE CH CY DE DK ES FI FF GB GR IE IT LI LU MC NL PT SE
 AET WO 9934208 A1 WO 1998-US27865 19981230; AU 9919503 A AU 1999-19503
 19981230; EP 1042669 A1 EP 1998-164342 19981230, WO 1998-US27865 19981230
 FDT AU 9919503 A Based on WO 9934208; EP 1042669 A1 Based on WO 9934208
 PFAI US 1997-1647 19971231
 IC ICM G01N033-49
 ICS G01N033-86; **G06F019-00**
 AB WO 9934208 A UPAB: 19990922
 NOVELTY - Predicting the presence of an abnormal level of at least 1 **protein** in the clotting cascade from at least one time-dependent measurement **profile** involves using a model that represents the relationship between the abnormal level of the **protein** in the clotting cascade and a set of predictor variables.
 DETAILED DESCRIPTION - Predicting the presence of an abnormal level of at least 1 **protein** in the clotting cascade from at least one time-dependent measurement **profile** comprises:
 (1) performing at least one time-dependent measurement on an unknown sample and measuring a respective property over time to derive a time-dependent measurement **profile**;
 (2) defining a set of predictor variables which define the data of the time-dependent measurement **profile**;
 (3) deriving a model that represents the relationship between the abnormal level of the **protein** in the clotting cascade and the set of predictor variables and
 (4) using the model of step (c) to predict the existence of the abnormal level of the **protein** in the clotting cascade and to predict which **protein** or **proteins** in the clotting cascade are at an abnormal level.
 The prediction of the **protein** or **proteins** at an abnormal level is a better prediction than an abnormal clot time alone.
 An INDEPENDENT CLAIM is also included for presenting a relationship between data from an assay relating to thrombosis-hemostasis on an unknown sample and data from assays relating to thrombosis-hemostasis from known sample populations which comprises:
 (A) providing data from at least one time dependent measurement **profile** for each of known blood samples;
 (B) performing at least one time-dependent measurement on an unknown blood sample and measuring a respective property over time to derive at least one time-dependent measurement **profile** for the unknown blood sample;
 (C) transforming data from step (2) to predictor variables which capture the information content of both the unknown blood sample time-dependent measurement **profile** and the known blood sample time-dependent measurement **profiles** and
 (D) presenting the data from the unknown blood sample time-dependent measurement **profile** relative to the data from the known blood sample time-dependent measurement **profiles**.

USE - Used for estimating the concentration of at least one **protein** in the clotting cascade.

Dwg.0/30

FS CPI EPI

FA AR; DCN

MC CPI: B04-B04D2; B04-N04; B11-C07; B12-K04

EPI: S03-E14H; S03-E14H1; T01-J

TECH UPTX: 19990912

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preferred method: In the prediction of the **protein** at an abnormal level, the specificity is at least 0.95 and the sensitivity is greater than 0.6. The samples with a measured concentration of less than 50% of normal for a specific factor are defined as being at an abnormal level.

The time-dependent measurement **profile** is at least one optical **profile** is produced by an automated analyzer for thrombosis and hemostasis testing. A number of optical measurements at one or more wavelengths are taken over time to derive at least one optical **profile**. The optical measurements correspond to changes in light scattering and/or light absorption in the unknown sample. The optical measurements are taken over time to derive at least one optical **profile**, and the optical measurements are each normalized to a first optical measurement. At least one optical **profile** is produced automatically by an analyzer. The unknown sample is automatically removed by an automated probe from a sample container to a test well. One or more reagents are automatically added to the test well to initiate the property changes within the sample. The development of the property over time is automatically optically monitored.

A predicted congenital or acquired imbalance or therapeutic condition is automatically stored in a memory of the automated analyzer and/or displayed on the automated analyzer. Assays for confirming the existence of the congenital or acquired imbalance or therapeutic condition are automatically effected. A set of data from known samples is produced which is used for deriving the model. The data from known samples is obtained by performing assays on the known samples.

Time dependent measurement **profiles** include at least two **profiles** from assays initiated with PT reagents, APIT reagents, fibrinogen reagents and TT reagents.

In (1), steps (3) and (4) comprise transforming a set of input parameters from the time-dependent measurement **profiles** for the known blood samples and the unknown blood sample, to corresponding individual output neurons whose location on an output map corresponds to the respective input data. Step (4) also comprises:

- (1) selecting weight vectors;
- (2) selecting a sample from a training set;
- (3) identifying best matching winning neuron at a particular time;
- (4) updating weight vectors and

(5) repeating steps (1)-(4) until the map reaches equilibrium.

In step (3), data from the time-dependent measurement **profiles** is transformed into predictor variables that characterize timing, rate and magnitude of changes during the time-dependent measurement **profile** and the predictor variables are used as input for neural networks. The definition of the predictor variables is a position in a self-organizing feature map, trained with data from the time-dependent measurement **profiles** for the known blood samples.

Preferred materials: The **proteins** comprise factors II, V, VII, VIII, IX, X, XI and/or XIII. The known blood samples and the unknown blood sample are samples of whole blood, plasma, or other part of whole blood. The known blood samples are samples of which are information is known relating to one or more intrinsic or extrinsic clotting factors and/or therapeutic agents. The known blood samples are samples of which are known the presence or absence of one or more abnormalities relating to at least one of fibrinogen level, oral anticoagulant, heparin, and one or more factor levels.

In step (4) one or more of normal sample, presence of heparin, and one or more factor deficiencies are presented on a PT map, or at least one of normal specimen, presence of heparin, abnormal fibrinogen, oral

anticoagulant, and one or more factor deficiencies are presented on an APTT map. The predictor variables are in terms of a standard deviation from a mean of at least one known blood sample population, and the unknown blood sample is characterized by variation from the mean of the known blood samples for each predictor variable.

L123 ANSWER 46 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 1993-418399 [35] WPIX
 DNN N1993-21236 DMC 01993-121896
 TI Calculating relative stabilities of two molecules from conformational free energies.
 SC A11 EPI R04 D15 D16 R17 T01
 IN PHILIPPINE, I
 PA (UTCO) UNIV COLUMBIA NEW YORK
 CYC 21
 PI WO 9917221 A1 19990405 (1999-5)* EN 83p G06F017-16 <--
 FW: AT BE CH CY DE DK ES FI FR GB GF IE IT LU MC NL PT SE
 W: AU CA JP MX US
 AU 9297796 A 19930424 (1993-5)
 US 6178384 B1 20010123 (2001-7) G06F019-00 <--
 ADT WO 9917221 A1 WO 1998-020368 19980929; AU 9297796 A AU 1998-37796
 19980929; US 6178384 B1 US 1997-940110 19970929
 FDT AU 9297796 A Pased on WO 9917221
 PRAI US 1997-940146 19970929
 IC PCT G06F017-16; G06F019-00
 AB WO 9917221 A UPAB: 19990908
 NOVELTY - The relative stabilities of two molecules, having defined connectivities and existing in defined environments, are determined by computer analysis of data representing low-energy minimum conformations to generate a total configuration integral (TCI) and thus the conformational free energy, (CFE) for both molecules, then comparing the CFEs.

DETAILED DESCRIPTION - The initial data set is derived from molecular connectivity, a potential energy **function** and a conformational search method. A configuration integral, in all degrees of freedom, is calculated based on the contribution of each molecular conformation by performing importance-sampling, multidimensional Monte Carlo (MC) integration over a multidimensional volume (dependent on harmonic vibrational frequencies of the conformation and temperature) encoding a current conformation of the molecule in conformational space. This process is repeated for every conformation in the data set and the contributions summed to generate TCI.

The importance sampling involves preferentially sampling regions of the volume having dominant contributions to the CFE of the conformation and uses a physical background of normal modes of vibration, with sampling performed using an atomic coordinate transform based on eigenvectors of the Hessian matrix (H_i) associated with the particular conformation.

INDEPENDENT CLAIMS are also included for the following:

(1) similar method for analyzing molecular complexes to determine relative binding affinities of molecules for a given host molecule; and
 (2) apparatus for the analysis.

USE - The method is used to select drug candidates.

ADVANTAGE - The method should identify agents with greater specificity for, and activity at, target sites. It does not require expensive free energy simulations or computational 'alchemy'.

Dwg.0/4

FS CPI EPI
 FA AP; DCN
 MC CPI: A03-A; B04-E01; B04-N01; B11-C09; B12-K04; D05-H09
 EPI: T01-E04; T01-J04C

TECH UPTX: 19990902

TECHNOLOGY FOCUS - ORGANIC CHEMISTRY - Preferred method: The low-energy minimum conformations are 10-35 kJ/mole above the lowest energy conformation and the potential energy **function** is based on molecular mechanics or on ab initio, semi-empirical or density-functional quantum mechanics. The atomic co-ordinates used are external (Cartesian) or internal (bond lengths, bond or torsional angles)

and H_i is exact or an approximation. The predetermined environment is presence of solvent, vacuum or a third molecule, and the molecules may be subjected to extrinsic environmental forces. The molecules being analyzed can be cyclic or acyclic, particularly water, biopolymers, bio-oligomers, **proteins**, enzymes, **peptides**, nucleic acid, carbohydrates, glycoproteins, receptors, lipid bilayer, biocompatible polymers, metals, inorganic compounds, pharmaceuticals, drugs and drug candidates.

To analyze molecular complexes, four data sets are generated: one for each of the two test molecules and one for each of their complexes. The CSE difference between:

(1) first and second test compounds (Δ_1 and Δ_2) are calculated, and the total difference (Δt) defined as $\Delta_2 - \Delta_1$.

If $\Delta\epsilon$ is negative, then the first test compound has the higher affinity, if it is positive then the second compound has higher affinity. The procedure may be repeated with third molecules and complexes. In a particular case, the method is used for different stereoisomers of the same compound and in that case only the difference $\Delta\epsilon$ needs to be calculated. In this process, internal atomic coordinates may also be relative spatial position of the components of the complex, defined by 3 translational and 3 rotational degrees of freedom.

TECHNOLOGY FOCUS - COMPUTING AND CONTROL - The apparatus comprises devices able to generate the data sets; determine CFE and calculate differences between CFEs. Any workstation or supercomputer should be suitable, running standard software such as the 'Cyclop' program or its equivalents.

Mathematics: The method uses a new MC integration technique, designated mode integration (MINTA), that operates on all the degrees of freedom, with the harmonic approximation used to provide highly accurate numerical integration of the high-dimensional configuration integral.

A sampling **function**, very similar to the partition **function** near the low energy minimal on the potential energy surface, is derived by applying the harmonic approximation to each low-energy well by a local, second-order Taylor expansion of the potential energy **function**. The sampling **functions** are normalized gaussians defined in terms of the bottom of the particular energy well and its associated local H_i .

L123 ANSWER 37 OF 73 WFMX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 1999-326687 [27] WFMX
 INN 01999-245026 DDC C1999-096626
 TI Derivatives of alpha-conotoxin and their analogs.
 DC B04 D16 S03 T01
 IN CARTIER, G E; KOEPFER, S C; MCINTOSH, J M; OLIVEIRA, B M; RIVIER, J E;
 SHEN, G S; SHOU, K; YOSHIKAMI, T
 PA (COGNITIX) COGNETIX INC; (SALK) SALK INST; (UYCA-N) UNIV CASE WESTERN
 RESERVE; (UTAH) UNIV UTAH FES FUND
 CYC 82
 PI WO 9921878 A1 19990506 (199907) * EN 175p COVFO07-08 ***
 EW: AT BE CH CY DE DK EA ES FI FF GB GH GM GR IE IT KE LS LU MC MW NL
 OA PT SD SE SC UG ZW
 W: AL AM AT AU AS BA BE BG BP BY CA CH CN CQ CC EE DK EE ES FI GB GE
 GH GM HE HI ID IL IS JP KE KG PF KP KG LC LV LR LS LT LU LV MD MG
 MK MR MW MX NL NE FI FT FG FU TO SE SF SI SK SL TJ TN TH TT JA UG
 UE VN YU ZW
 AU 9911143 A 19990517 (199909) COVFO07-08 ***
 EP 1031588 A1 19990906 (200004) EN COVFO07-08 ***
 E: AT BE CH CY DE DK ES FI FR GB GR IE IT LI LV MC NL PT SE
 ADT WO 9921878 A1 WO 1999-0321-68 19991023; AU 9911143 A AU 1999-11143
 19991023; EP 1031588 A1 EP 1998-053868 19981023, WO 1998-053868 19981023
 FCT AU 9911143 A Based on WO 9921878; EP 1031588 A1 Based on WO 9921878
 PRAI US 1997-65814 19971114; US 1997-62783 19971024
 IC ICM C07K007-08
 ICS G01N014-08; G01N033-68; G06F017-30; G06F019-00
 AB WO 9921878 A UPAB: 19990714

NOVELTY - Derivatives (I) of alpha-conotoxin MII (II) with practically the same activity as (II) are new.

DETAILED DESCRIPTION - (I) are of formula X-Cys-Cys-X-X1-X2-X-Cys-X3-X-X4-X5-X-X-Cys (I')

X, X1, X2, X3 and X4 = same or different natural, modified or non-natural amino acids;

X5 = His or Asn

INDEPENDENT CLAIMS are also included for the following:

(1) a peptide analog, peptide mimic or mimetic

(A) of (II) specific for one or more subtypes of neuronal nicotinic acetylcholine receptor (nAChR);

(a) screening compounds for antagonistic activity at an nAChR subtype by comparing on and off rates for (II) and a test compound to the different subunits of nAChR;

(b) derivative, peptide analog, peptide mimetic or mimetic (P) of (II) which selectively modulates biological activity of nAChRs;

(4) method for determining the three-dimensional (3-D) structure of an alpha-conotoxin analog (IIa), or part of it;

(5) method for selecting and designing compounds able to bind to nAChR by analyzing the 3-D structure of (II), or part of it; and

(6) machine-readable data storage medium which can display a graphical 3-D representation of an MII peptide or of an alpha-conotoxin (or parts of them).

ACTIVITY - Antitumor; anti-smoking; cardiovascular.

MECHANISM OF ACTION - Specific interaction with neuronal nicotinic acetylcholine receptors.

USE - (I), and its mimetics, are useful as cardiovascular agents; for treating or diagnosing small-cell lung carcinoma; and as gastric motility, urinary incontinence and anti-smoking agents.

ADVANTAGE - (I) and their mimetics can be designed to be selective for particular subtypes of nAChR, particularly the alpha 3 beta 2 and alpha 3 beta 4 subtypes.

Dwg.0/14

FS CPI EPI

FA AB; DCN

MC CFI: B04-C01D; B04-H01; B04-N04A; B11-C08E; B12-K04E;
B14-F01; B14-H01; B14-M01B; B14-N09; D05-H09; D05-H12B2

EPI: S03-E07; S03-E14H; T01-J; T01-J05B

TECH UPTX: 1990714

TECHNOLOGY FOCUS - PHARMACEUTICALS - Preferred compounds: (I) is C-amidated and has X1 = Asn; X2 = Pro or hydroxy-Pro; X3 and X5 = His; X4 = Glu. They have practically the same biological activity as (II). (A) are specific for one of the receptor subtypes alpha3beta2; alpha3beta4 or alpha2beta2.

TECHNOLOGY FOCUS - ORGANIC CHEMISTRY - Preferred process: To determine the 3-D structure of (IIa), NMR data are obtained for (IIa) and analyzed, with reference to known structural co-ordinates for MII, by molecular replacement. Selected (IIa) may then be tested for binding to an nAChR; inhibition of binding of ligand to nAChR and/or inhibition of biological function mediated by a natural ligand of nAChR. To select a compound able to bind to nAChR, the 3-D structure of MII is analyzed to identify regions that interact with particular functional groups, then databases examined for compounds having structures that best fit these interacting regions. To design a compound able to bind to nAChR, a 3-D representation based on co-ordinates defining the structure (of part) of (II) is displayed and interaction with parts of a known ligand are characterized to identify suitable groups for replacement. Groups that can replace one or more parts of the ligand, while retaining at least some binding affinity, are then identified from a knowledge base of possible substitution groups. Alternatively, positions that are preferred with respect to favorability of interactions with functional groups are identified in the 3-D structure, and parts of a known ligand close to these positions are characterized. From a knowledge base, fragments or molecules that permit connection of the preferred positions to portions of the ligand are identified and the structure of the ligand is modified by

covalent attachment of these fragments or molecules, in the proper orientation and location. Compounds designed this way are then tested for modulatory activity at nAChRs.

Preparation: (I) are produced by usual methods of **peptide** synthesis.

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preparation: (I) may be produced by usual recombinant DNA methods (not exemplified).

TECHNOLOGY FOCUS - COMPUTING AND CONTROL - The data storage medium displays a 3-D representation, based on all or some of the atomic coordinates taken from a table, reproduced in the specification, of the 3122 atoms of (II), deposited at the Brookhaven **Protein** Data Bank as 1m2c, or on coordinates with a root-mean-square deviation, for conserved backbone **protein** atoms, of not over 1.8Angstrom. The medium may carry two sets of data: one, a Fourier transform of at least part of the coordinates specified above and the other the coordinates of a molecule or molecular complex.

L123 ANSWER SP OF 7/3 WPIX COPYRIGHT 2001 DEFVENT INFORMATION LTD
 AN 1998-326593 (27) WPIX
 DNN N1 199-144371 P00 C1994-09-074
 TI Voltage-gated, pH-sensitive potassium channel useful in gene therapy.
 DC P04 16 C03 T01
 IN FALCOFF, L; SCHREIBER, M; SILVIA, C
 PA (MMW) MMW WASHINGTON
 CYC 64
 PI WO 9920754 A1 1998-04-09 (19980510) EN C12N015-11
 RW: AT BE CH CY DE ES FI FR GE IM CR IE IT KE LS LJ MC MW NL
 OA PT RU SP SU TG SW
 W: AL AM AT AP AL BA BB BC BE BI CA CH CN CU CS DE DF EE ES FI GE GD
 GE GH GM HE HI ID IL IS JE KE PG PR KZ LC LK LR LS LT LT LV MD
 MG MK MN MW MX NO NJ PL PT RU SD SE SG SI SK SL TC TM TR TT UA
 US UK UC VI YU SW
 AU 9911122 A 19990510 (199905) C12N015-11
 EP 1029042 A1 20000831 (2000041) EN C12N015-11
 E: AT BE CH CY DE DK ES FI FR GE IE IT LI LJ MC NL PT SE
 ADT WO 9920754 A1 WO 1998-US22321 19981021; AU 9911122 A1 1999-11122
 19990511; EP 1029042 A1 EP 1998-95185 19981021, WO 1998-US22321 19981021
 FDT AU 9911122 A Based on WO 9920754; EP 1029042 A1 Based on WO 9920754
 PRAI US 1448-76172 199-0217; US 1397-63138 13971022
 IC ICM C12N015-11
 ICS C07H021-04; C07K014-00; C07K016-00; C12N015-65;
 C12N015-85; C12Q001-00; C12Q001-68; G01N033-55; G01N033-567;
 G06F019-00
 AB WO 9920754 A UPAK: 19990714
 NOVELTY - A voltage-gated, pH sensitive potassium channel Slo3, expressed in spermatocytes, is new.
 DETAILED DESCRIPTION - Slo3 has, as a monomer, calculated molecular weight 130-156 kD; has unit conductance (as a **functional** tetramer, when expressed in Xenopus oocytes) of 60-120 pS; has increased activity at intracellular pH above about 7.1 and binds specifically to polyclonal antibodies against sequences (P1), (P2), (P3) or (P4), all given in the specification, with 1113 and about 110, 1056 and 1020 amino acids, respectively.
 INDEPENDENT CLAIMS are also included for the following:
 (1) isolated nucleic acid (I) encoding Slo3;
 (2) isolated nucleic acids (Ia) encoding at least 15 contiguous amino acids from Slo3, and their conservatively modified variants;
 (3) antibodies (Ab) that bind selectively to murine or human Slo3;
 (4) expression vector containing (I);
 (5) host cell transfected with this vector;
 (6) method for identifying agents (A) that increase or decrease ion-flux through a pH-sensitive potassium channel;
 (7) detecting Slo3 in mammalian tissue by reaction with a selective binding agent;

(3) computer method of screening for mutations in Slo3 genes; and
 (3) computer method for identifying a three-dimensional Slo3 structure.

ACTIVITY - Contraceptive; fertility-regulating.

MECHANISM OF ACTION - Slo3 is involved in sperm capacitation and/or the acrosome reaction, essential steps in fertilization.

USE - Slo3, and (I), encoding it, are used to identify specific inhibitors and activators (potentially useful for treating infertility and as contraceptives), also for studying sperm physiology in vitro. Slo3-specific antibodies are used for diagnostic detection of Slo3 expression. Slo3, as part of a chimera with another channel protein, can be used as a reporter for measuring changes in potassium concentration, current flow, ion flux, etc.

Fragments of (I) are useful as probes for identifying homologs, variants and mutants associated with disease; to detect Slo3-related mRNA or protein; for chromosomal localization; in gene therapy; for identifying potential modulators; to measure up-regulation of Slo3 in drug screening; assays and for production of recombinant Slo3 protein.

Dwg.0/4

FS CPI EPI

FA AB; ICH

MC CPI: E04-E01; E04-E02; E04-E01; E04-H01; E11-C07A; B12-K04;

D05-H09; D05-H11; D05-H12A; D05-H12E; D05-H14; D05-H17A2

EPI: S03-E14H4; T01-J

TECH UPTM: 19990714

TECHNOLOGY FOCUS - PHYSIOLOGY - Preferred nucleic acid: (I) encodes murine or human Slo3, specifically (P1), murine, or (P3) or (P4), both human. (I) is hybridizes under moderately stringent conditions with sequences (N1), (N2), (N3) or (N4), of about 0.7, 0.4, 0.2 and 0.1 kb, respectively, given in the specification.

Primer pairs (N5-N12) suitable for amplification of these four sequences are also given in the specification. Alternatively, (I) encodes the monomer of a pH-sensitive potassium channel that has a core domain with more than 60% amino acid sequence identity with the 35-641 region of the Slo3 core domain and is recognized by antibodies raised against the core domain of (P1), (P2), (P3) or (P4).

CTCGAACTCCCTAAAATCTTACAGAT (N5)

TTCCGTTGAGCCAGGGTCACCAAGAATT (N6)

TCTGCTTGTGAAGCTAAATCT (N7)

TTTCAAAGCTCTTTAGCGGTAA (N8)

TTATGCCTGGATCTGCACTCTACATG (N9)

ATAGTTCTCTCTACTAACCGAAA (N10)

GGCAGCGCTCATTTCTCTCCTT (N11)

TGCCCAAAACCTCAACCCAAAATA (N12)

Preferred process: To identify (A), a test compound is applied to a eukaryotic cells (or its membrane) that has expressed a Slo3 monomer, and any effect of the compound determined, particularly by measuring the whole cell conductance to detect increase/decrease in ion flux. Particularly the monomer is a recombinant protein. To detect Slo3, the specific binding agent is an antibody, oligonucleotide or nucleic acid probe. To screen for mutations in Slo3 genes, sequences (particularly where associated with disease) are compared, over at least 30 nucleotides (nt), with sequences (N1), (N2), (N3) or (N4).

Preparation: Antibodies against Slo3 are raised by usual immunization and cell fusion methods.

TECHNOLOGY FOCUS - COMPUTING AND CONTROL - To determine the three-dimensional structure of Slo3 protein, data covering at least 10 amino acids of the protein, or 30 nt of the nucleic acid, are analyzed to generate a primary protein structure. This may then be converted, using energy terms associated with the structure, to secondary or tertiary structures, and optionally anisotropy terms in the tertiary structure used to devise a quaternary structure. The resulting structures may be used to identify ligand-binding sites and, from these, the ligands themselves.

TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preparation: Expressed sequence tag (EST) databases were searched with the C-terminal region of murine Slc1 to identify the homolog AAQ2536 (GenBank), derived from a mouse promyelocytic WEHI-3 cell line cDNA library. A 1254 bp amplicon, generated from the plasmid containing this EST, was used to screen a mouse testis cDNA library to isolate the full-length sequence (NL), from two overlapping fragments. The murine sequence was used to provide polymerase chain reaction primers for amplification of human testis cDNA, resulting in sequence- (NL). This was used to probe the same library under moderately stringent conditions, to isolate 6 clones, encoding the two variants (P3) and (P4). Once isolated, the cDNA may be expressed in standard vector/host cell systems.

L123 ANSWER 39 OF 41: WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 1999-200002 (14) WPIX
 DNN NL 990-100426 DDC C12Q01-63
 TI Molecular design by exon shuffling within a library of DNA sequences.
 DC P04 D16 T01 T06
 IN KARUBE, I; OKABE, Y; SUMIKURA, E
 PA (KAFU-1); KARUBE I
 CYC 81
 PI WO 9911618 A1 19990311 (199919)* JA 42p C12Q01-63
 RN: AT BE CH CY DE EW RA RS FI FR GP GH GM GP IE IT KE LS LU MC MW NL
 'A PT UD SE SU UG ZW
 W: AL AM AT AU AC RA B6 B9 BR B7 CA CH CN CU CZ DE DK EE EG FI GB GE
 GH GM HR HQ ID IL IS JE KE KG KF KZ LC LK LR LS LT LU LV MD MG MK
 'B MW NX NO NZ PL PT EG RU UD SE SG SI SK SI TJ TM TR TT UA UG US
 SU VN YU ZW
 AU 9888667 A 19990312 (199921) C12Q01-63
 ADT WO 9911618 A1 WO 1999-030174 19980126; AU 9888667 A AU 1998-08867 19980626
 FDT A9 990626 7 A Based on WO 9911618
 PRAI UP 1997-040679 19970826
 IC ICM C12Q01-63
 ICS C07K021-00; C07K001-00; C07K007-00;
 C07K014-00; C12N015-11; G06F015-42
 AB WO 9911618 A UPAB: 19990518
 NOVELTY - Molecular design is carried out by shuffling exons within a library of similar DNA sequences, followed by ranking of the products of reshuffling by their fit to the desired **functionality**, and repeating the shuffling and ranking until a product of the desired **functionality** is obtained. A similar process may be carried out using **polypeptide** sequences.
 DETAILED DESCRIPTION - New **functionality** is obtained or existing **functionality** is improved in **polypeptides** or nucleic acids by:
 (1) synthesising a group of **polypeptides** or nucleic acids with differing sequences;
 (2) measuring the fit of these sequences to the desired **functionality** by laboratory experiment;
 (3) ranking the sequences in order of this degree of fit;
 (4) preparing a 'shuffling library' of sequences by shuffling parts of these sequences between individuals ranking highly;
 (5) synthesising the resulting shuffled sequences;
 (6) repeating the above steps a sufficient number of times to obtain one or more sequences with close fit to the desired **functionality**
 USE - The technique can be used for the improvement of **function**, and devising of new **functions**, of **polypeptides** and nucleic acids, for a broad range of uses such as drugs and foodstuffs.
 ADVANTAGE - **Evolution** of high **functional** fit is rapid compared to the use of conventional genetic **evolution** algorithms.
 Dwg.3/4
 FS CPI EPI
 FA AB; DCN

MC CPI: B04-C01; B04-E01; **B04-N0400E**; B11-C01; **B12-K04E**;
D05-H

EPI: T01-X; T06-D02; T06-D16

TECH UPTX: 19399510

TECHNOLOGY FOCUS - BIOTECHNOLOGY - The shuffling of parts of the sequences may be supplemented by steps in which one or more bases or amino acids of the sequence are mutated (by addition, deletion or substitution).

L123 ANSWER 40 OF 72 WPIX COPYRIGHT 2001 LEPWENT INFORMATION LTD
AN 1999-15935 [14] WPIX
LBNR N1001-111670 EPO 00-04118
TI New *Streptococcus pneumoniae* Response Regulator (RR) **polypeptide** and polynucleotide - Useful as diagnostic reagents and for prevention and treatment of *Streptococcus* infections which cause conjunctivitis, sinusitis and meningitis.
DC B04 C16 E01
IN BISWAS, S; GE, Y; HOLMES, D J; INGRAHAM, K A; THROUP, J; WALLIS, N G; ZAHARIAH, M; HOLMES, D; INGRAHAM, K; WALLIS, N
PA (SWIK) SMITHKLINE BEECHAM CORP
CYC 27
PI EP 900846 A2 19990310 (199914) EN 56p C12N015-31
F: AL AT BE CH CY DE DK ES FI FR GE GR IE IT LT LU LV MC MK NL PT
SD SE SI
CA 2143656 A1 19990309 (199944) EN C12N015-31
JP 11196238 A 19990721 (199950) 56p 97K014-315 --
ABT EP 900846 A2 EP 1998-367054 19980301; CA 2243656 A1 CA 1998-2243656
19980303; JP 11196238 A JP 1998-194006 19980909
PRAI US 1997-60714 19970909
IC ICM C07K014-315; C12N015-31
ICU A61P031-00; A61K031-70; A61F037-00; A61F038-10; A61K039-045;
A61P039-09; A61K039-305; A61K 41-00; A61F043-00; C07K014-705
; C07K016-12; C12N015-04; C12P021-00; C12P021-08;
G01N033-15; G01N033-48; G01N033-50; G01N033-55; G01N033-566;
G01N033-569; G01N033-577; G06F019-00; G11B023-00
ICI C12N015-09, C12R001:46
AB EP 900846 A1 EPAB: 19990413
NOVELTY - An isolated **polypeptide** comprising an amino acid at least 70% identical to (I), a fully defined 428 *Streptococcus pneumoniae* Response Regulator (RR) **protein**, which is part of the two component signal transduction system (TCSTS).
DETAILED DESCRIPTION - INDEPENDENT CLAIMS are included for the following: (1) an isolated polynucleotide (II) comprising a nucleotide sequence at least 70% identical to a polynucleotide encoding (I); (2) an expression system comprising polynucleotide (II); (3) a host cell comprising the expression system or a membrane of; (4) an antibody comprising the expression system or a membrane of; (5) an agonist or antagonist (III) to (I); (6) use immunospecific for (I); (7) an agonist or antagonist (III); (8) polynucleotide (II); or (iii) of: (i) an agonist or antagonist (III); (ii) polynucleotide (II); or (iii) a nucleic acid molecule that modulates expression of (II); in therapy; (7) preparation of (I); (8) an isolated polynucleotide comprising a nucleotide sequence at least 70% identical to (V), a fully defined 3453 bp nucleic acid given in the specification; and (9) a **polypeptide** which comprises an amino acid sequence at least 70% identical to (IV), a fully defined 428 amino acid **protein** given in the specification.
USE - RR sequences are useful for diagnosing a disease or susceptibility to a disease related to RR **polypeptide** (I) levels by determining the presence/absence of a mutation in the gene, or analysing for the presence or amount of (I) expressed in a sample, due to an infection with an organism with the gene claimed. They can diagnose the state and type of infection. They are also useful for screening for compounds (III) which stimulate or inhibit **polypeptide** (I) **function** by determining the effect on activity of (I) (claimed). Agonists and antagonists are useful for treatment of conditions associated with RR imbalance, and are therefore potential antibacterial compounds. RR polynucleotides are useful for genetic immunisation and antisense sequences are useful for inhibition of expression of (II). RR **polypeptides** are administered to treat individuals in need of RR

polypeptides (directly or via a vector i.e. gene therapy), and as an antigen for inducing an immune response. They can prevent adhesion of bacteria to matrix **proteins**, and are useful for use on wounds and body implants to prevent bacterial infection. Anti-**RF** antibodies induced by the **polypeptide** are useful for preventing or treating infections, especially bacterial infections, and also for isolating clones expressing **RF** (I), or purifying the **polypeptide** by affinity chromatography. Diseases prevented, diagnosed and treated include those caused by bacterial infection, especially *Streptococcus pneumoniae* infections, which cause otitis media, conjunctivitis, pneumonia, bacteraemia, sinusitis, pleural empyema, endocarditis and especially meningitis.

Dwg.0/0

FS CPI EPI

FA AB

MC CPI: B04-E01F; B04-E06; B04-F010E; B04-G07; B04-G0700E; B04-G21; B04-G2100E; B04-G22; B04-G2200E; B04-N03A0E; **B12-K04A4**; **B12-K04E**; **B12-K04F**; B14-A01; B14-A01B2; B14-F01; B14-F02; B14-K01; B14-L01; B14-L06; B14-N02; B14-N03; B14-N04; B14-N16; B14-N17B; B14-S03; B14-S11B; D05-C12; D05-H04; D05-H07; **D05-H09**; D05-H11A1; D05-H11A2; D05-H11B; D05-H12A; D05-H12D2; D05-H14; D05-H17A6

EPI: S03-E14H4

L123 ANSWER 41 OF 73 WPIW COPYRIGHT 2001 DEFENT INFORMATION LTD

AN 1999-120378 [1C] WPIX

DNN N1999-037830 EPO C1999-035124

TI Inferring the biological **function** of a **protein** using a ligand database - from the **function** of known **proteins** to which a ligand binds which is also capable of binding to the **protein** under examination.

DC B04 T01

IN IMAMURA, M; ITAI, A

PA (ITAI-I) ITAI A

CYC 83

PI WO 9901409 A1 19990114 (199910)* JA 33P C07B061-00
 FW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL
 OA PT SD SE SZ UG ZW
 W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE
 GH GM GW HR NU ID IL IS JP KE KG KR KZ LC LK LF LS LT LU LV MD MG
 MK MN MW MK NO NC PL PT RO PU SD SE SG SI SK SL TJ TM TR TT UA UG
 US UZ VN YU SW

AU 9879366 A 19990125 (199923) C07B061-00

EP 1008572 A1 20000614 (200033) EN C07B061-00

R: CH DE FR GB IT LI

AET WO 9901409 A1 WO 1998-JP2986 19980702; AU 9879366 A A1 1998-79366

19980702; EP 1008572 A1 EP 1998-929306 19980702, WO 1998-JP2986 19980702

PDT AU 9879366 A Based on WO 9901409; EP 1008572 A1 Based on WO 9901409

PFAI JP 1997-177933 19970703

IC ICM C07B061-00

ICS C07K001-00; G06F015-40; G06F017-30;

G06F017-50

AB WO 9901409 A UPAB: 19990310

Inferring the biological **function** or activity of a **protein** whose three-dimensional configuration is known or can be inferred comprises: (i) constructing a database of three-dimensional structures of a large number of physiologically active compounds which are known to bind to target **proteins** with known biological **function** or activity; (ii) extracting candidate ligands to the **protein** (I) under examination from the database and their ability to form a complex with (I) is investigated with a suitable docking program; (iii) a candidate ligand is chosen from the database which has the best docking with (I), i.e. forms the most stable complex with (I); and (iv) it is then inferred that the **function** or activity of (I) is similar to that of the target **protein** or **proteins** recorded in the database as binding to the chosen ligand. A suitable

program for carrying out this investigation is ADAM&EVE (described in WO9613785, Akiko Itai) which determines the best docking conformation of the ligand with the target **protein** and then calculates the energy of interaction between them in this conformation.

USE - The method is rapid and accurately infers the biological function or activity of **proteins**, such as **proteins** of previously unknown function discovered during the human **genome** project, for use (for example) in metabolic investigations or drug design.

Seq.6/2

FS CPI EPI

FA AB

MC CPI: B01-A01; B01-A; B04-A-4; B04-N04; B06-I-01; B11-A17;
B10-S01; B11-C03E; B12-K04
EPI: P01-J06P; P01-J15

L123 ANSWER 42 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
AN 1999-108349 [10] WPIX

DNC C1999-032512

TI New **SecA polypeptides** and polynucleotides - useful as diagnostic reagents and for prevention and treatment of *Streptococcus pneumoniae* infections, especially meningitis.

DC E04 B10 S03

IN JAWORSKI, B D; O'DWYER, K M; SHILLING, L F; TRAINI, C M; WANG, M; WILDING, E J; QUINTE, K M

FA (SMIR) SMITHKLINE BEECHAM CORP; (SMIR) SMITHKLINE BEECHAM PLC

CYC 27

EP 894857 A1 19900203 (199010)* EN 34p C12N015-31
R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI PT LU LV MC MK NL PT
RO SE SI

CA 2238672 A 1998-01-1998-01 C12N015-31

JP 11192091 A 1999-0721 (19990721) 83p C12N015-09

ALT EP 894857 A2 EP 1998-09537; 19960713; CA 2238672 A CA 1998-2238672
19980724; JP 11192091 A JP 1998-250289 19980731

PFAI US 1998-099720 19980313; US 1997-54563 19970801

IC ICM C12N015-09; C12N015-31

ICS A61F031-70; A61K035-70; A61K038-00; A61K038-16; A61F039-09;
A61K039-395; A61K045-00; A61K048-00; C07K014-315;

C07K016-12; C12N001-15; C12N001-19; C12N001-31; C12N005-10;
C12P021-02; C12Q001-02; C12Q001-68; G01N033-56; G01N033-569;
G06F019-00; G11B023-00

ICA C12P021-08

ICI C12N015-09, C12F001:44

AB EP 894857 A UFAB: 19990310

A **secA polypeptide** at least 70% identical to sequence (I), a fully defined 837 amino acid **protein** given in the specification is new. Also claimed are: (1) an isolated polynucleotide (II) (DNA or RNA) complementary or at least 70% identity to a polynucleotide encoding (I); (2) a vector comprising polynucleotide (II); (3) a host cell comprising the vector; (4) an antibody against **secA polypeptide** (I); (5) an antagonist which inhibits activity or expression of **secA polypeptide** (I); (6) an isolated polynucleotide (IV) comprising at least 70% identity to a polynucleotide encoding **polypeptide** (III), a fully defined 693 amino acid **protein** given in the specification; and (7) a computer readable medium stored with data selected from: (i) **secA** polynucleotides (II)-(IV) or **polypeptides** (I)/(III); (ii) a set of polynucleotides or **polypeptides**, where at least one sequence is a **secA** polynucleotide or **polypeptide**; and (iii) a data set representing **secA** polynucleotides or **polypeptides**.

USE - **SecA polynucleotides and polypeptides** are useful for diagnosing susceptibility to diseases by detecting mutations or polymorphisms in the **secA** gene or analysing for the presence of amount of **secA polypeptide** expressed in a patient sample (claimed). **SecA** PCR probes are useful for diagnosing diseases, and can characterise the response of the infectious organism to drugs. **SecA polypeptides**

and polynucleotides are also useful for screening for antagonists, agonists and drugs against infectious micro-organisms by binding with *secA* polypeptide (1) and observing interaction and activation or inhibition of the **polypeptide function** (claimed). *SecA* polypeptides and antagonists are useful for treating conditions associated with abnormal *secA* protein levels (claimed), and *secA* agonists, antagonists and drugs are bacteriostatic and bacteriocidal compounds which can be used in treatment to enhance (agonist) or block (antagonist or antisense sequence) *secA* activity, therefore treating microbial diseases, especially *Streptococcus pneumoniae* diseases including otitis media, otitis externa, tonsillitis, pneumonitis, sinusitis, pleural empyema, endocarditis and especially meningitis. Epitopes of *secA* polypeptides and polynucleotides are useful immunogens for polypeptides and polynucleotides are useful immunogens for prevention of bacterial infections, and proteins anti-*secA* antibodies for prevention of bacterial infections (gene *secA* polynucleotides can be used in genetic immunisation (gene therapy) using the vector to prevent infections (claimed)). *SecA* polypeptides, polynucleotides and their (anti)agonists can prevent adhesion of bacteria to matrix proteins, and are useful for use on wounds and body implants to prevent bacterial infection. *SecA* polypeptides and polynucleotides may also be used as reagents for differential screening methods e.g. using *secA* probes in RT-PCR to identify and quantify genes expressed in bacterial tissue. *SecA* polypeptides are useful for mapping genes to chromosomes, allowing gene inheritance to be studied through linkage analysis. The computer based method (7) is useful for performing homology identification by comparing a polynucleotide with *secA* sequences, and is also useful for comparing a polynucleotide assembly, by screening for overlapping sequences between a polynucleotide and *secA* polynucleotide at (7) (claimed).

FS CPI EPI
 FA AF
 MC CPI: B04-E03E; B04-E04E; B04-E05E; B04-J07; B14-G21; B04-J22; B04-N03A;
 B04-N03A0E; B12-K04A4; B12-K04E; B11-A01;
 B14-A01B2; B14-F01; B14-F02; B14-H01; B14-L06; B14-N02; B14-N03;
 B14-H04; B14-S03; D05-C12; D05-H07; D05-H09; D05-H11;
 D05-H12A; D05-H12D1; D05-H12E; D05-H14; D05-H17A6
 EPI: S03-E14H4

L123 ANSWER 43 OF 73 WPIX COPYRIGHT 2001 DEPARTMENT INFORMATION LTD
 AN 1999-062653 (06) WPIX
 DNN 111999-046539 DNC C1999-018641
 TI New isolated gicA1 **polypeptide** from *Streptococcus pneumoniae* - useful in diagnosis, treatment and prevention of bacterial infections.
 DC B04 D16 S03
 IN BURNHAM, M; FEDON, J C; JAWORSKI, D D; FALLENDER, H; LENKE, A L; PALMER, L
 M; WANG, M
 PA (SMIK) SMITHKLINE BEECHAM CORP; (SMIK) SMITHKLINE BEECHAM PLC
 CYC 37
 PI EP 889128 A2 19990107 (199906)* EN 44p C11N015-31
 F: AL AT BE CH CY DE DK ES FI FR GB GE IE IT LI LT LV MC MK NL PT
 FO SE SI
 CA 2236425 A 19990101 (199924) C11N015-31
 JP 11137263 A 19990526 (199921) 115p C11N015-09
 JP 2000210603 A 20000801 (200011) 40p C11N015-09
 ADT EP 889128 A2 EP 1999-105174 19990636; CA 2236425 A CA 1998-2236425
 1998-2236426; JP 11137263 A JP 1999-011843 19990701; JP 1000210603 A Div ex JP
 1998-2236426; 1998-223643 19990701; JP 1000210603 1999-0701
 PPAI US 1997-51079 13370701
 IC ICM C12N015-09; C11N015-31
 ICS A61K031-70; A61K031-7088; A61K031-10; A61K039-11; A61K039-395;
 A61K045-00; A61K045-00; A61F31-00; A61F31-04; C07K014-315
 ; C07K016-12; C12N001-10; C11N001-10; C11N001-11;
 C12N005-10; C12P021-01; C11N01-01; C01N131-15; C01N033-50;
 C01N033-53; C01N033-56; C01N033-561; C01N033-561; G06F017-30
 ; G06F019-00; C11B013-00
 ICA C11F021-04
 ICI C12N015-04; C12F001:46; C12F001:46; C12N015-04

AB EP 889129 A UPAB: 19390217

New isolate! **polypeptide** (II) comprise/s/is (at least 70 identity to) sequence (.), 687 amino acids (aa) or (4) 623 aa, over their entire length and/or is encoded by a recombinant polynucleotide comprising sequence (1) 2109 bp or (?) 1871 bp (all sequences fully defined in the specification). Also claimed are: (A) an isolated polynucleotide (I); (B) an antibody antigenic to or immunospecific for **polypeptide** (II); (C) a process for diagnosing or prognosis a (susceptibility to) disease in an individual related to expression or activity of (II); (D) a method for screening to identify compounds that activate or inhibit the **function** of (II); (E) an agonist/antagonist of the activity or expression of (II); (F) an expression system comprising (i) capable of producing (II) when present in compatibly host cell; (G) a host cell comprising the above expression system/membrane expressing (II); (H) a process for producing (II) comprising culturing the host cell; (I) a process for producing the host cell comprising the expression system for (II); (J) a recombinant host cell capable of expressing (II); (K) a computer readable medium with stored sequences (1)-(4); and (L) a computer based method for performing homology identification.

USE - (I), its agonists or (II) are used to treat conditions requiring increased activity or expression of (I) (conditions not cited), while conditions (particularly bacterial infections) requiring inhibition of (I) are treated by administering an antagonist, inhibitory nucleic acid or competitive **polypeptide** e.g. *S. pneumoniae* infection, particularly meningitis and also *Helicobacter pylori* infections e.g. related cancers, ulcers and gastritis. These antibacterial agents may also be used to treat in-dwelling devices to prevent infection or generally as wound treatments to prevent adhesion of bacteria to matrix **proteins**. (I)-related conditions, or susceptibility to them, can be diagnosed, staged or prognosis by (i) detecting mutations in (I)-encoding nucleic acid or (ii) by determining presence or amount of (I). (I), or cell membranes of (E), are used to screen for (II) (in any standard binding assay) and cells of (E) are used to produce recombinant (I), used (i) to raise Ab (for use in identifying/isolating (I)-expressing clones, for affinity purification, as therapeutic agent and in competitive drug screens); (ii) to identify (III) or specific receptors; (iii) in rational drug design and (iv) as immunogen for vaccines. (II), or its fragments, are used as antisense/ribozyme therapeutics; as probes and primers to isolate homologous sequences; to detect (mutant) (II); for chromosomal mapping; to determine bacterial serotype; for genetic immunisation; to screen for (III) and in rational drug design.

Dwg.070

FS CPI EPI

FA AB

MC CPI: B04-C01G; B04-E00F; B04-E05; B04-F0100E; B04-F1024; B04-G0700E;

B04-N04A; B11-C07A; B11-C08E; B12-K04A4; B14-A01;

B14-A11B2; B14-E08; B14-E10Z; B14-H01; B14-L01; B14-L06; B14-N17B;

B14-S11; D05-C11; D05-H07; **D05-H09; D05-H11; D05-H12A;**

D05-H11D; D05-H12E; D05-H14; D05-H17A

EPI: S03-E14H4

L123 ANSWER 44 OF 73 WPIX COPYRIGHT 2001 DEEPWENT INFORMATION LTD

AN 1999-034633 [03] WPIX

DNN NI 999-015941 DNC C1999-010385

TI Determination of 3D structure of **protein** - from amino acid sequence using hierarchical approach in which number of candidate structures decreases at each stage.DC B04-D16 S03 **T01**

IN DEBE, D A; GODDARD, W A

PA (CAL) CALIFORNIA INST OF TECHNOLOGY

CYC 63

PI WD 9848270 A1 19981029 (199903)* EN 72p GOING3-00

EW: AT BE CH CY DE DF EA ES FI FR GB HE GM GR IE IT KE LS LU MC MW NL OA PT SD SE SE UG ZW

W: AL AM AT AJ AS BA BB BG BR BY CA CH CN CY CZ DE DK EE ES FI GB GE GH GM GW HU ID IL IS JP KE KG KP KR HZ LC LK LS LT LU LV MD MG

ME MN MW MX NO NZ PL PT RU SD SE SG SI SK SL TJ TM TR TT UA UG
US VN YU SW

AU 9871466 A 19981112 (1998112) GO1N033-00
EP 977985 A1 20000209 (2000012) EN GO1N033-00
P: CH DE FR GB LI

ADT WO 9848270 A1 WO 1998-038077 19980421; AU 9871466 A AU 1998-71466
19980421; EP 977985 A1 EP 1998-018562 19980421, WO 1998-038077 19980421

FDT AU 9871466 A Based on WO 9848270; EP 977985 A1 Based on WO 9848270

PFAI US 1998-44124 19980420; US 1997-44124 19970422

IC ICM GO1N033-00
ICS G06F017-00

AB WO 9848270 A UPAB: 19981122
The three-dimensional backbone structure of an **n protein** is determined by selecting a finite set of torsion angles and generating an ensemble of conformations representing an exhaustive enumeration of self-avoiding backbone conformations for the selected set.

USE - The method is used for determining a **protein's** three-dimensional structure from its amino acid structure. This structure is often the key to elucidating the **function** and mechanism of the **protein** of interest, allowing researchers to discover new or more potent drugs.

ADVANTAGE - The method provides an accurate and easy to use method for an initial structure prediction.

Dwg.1/0

FS CPI EPI
FA AB; GI
MC CPI: B04-N04; B11-C08; B12-K04E; D05-H09;
E05-H18A
EPI: N03-E14; T01-C

L123 ANSWER 46 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 1999-018084 [02] WPIX
DNN N1999-014856 DNC C1999-005774

TI Determination of **protein** biological **function** - comprises use of amino acid sequences database containing the relevant information.

DC B04 T01
FA (IYAKU) IYAKU BUNSHI SEKKEI KENKYUSHO KK

CYC 1
PI JP 10287696 A 19981027 (199902) 11p C07K001-00 <--
ADT JP 10287696 A JP 1997-93577 19970411
PFAI JP 1997-93577 19970411
IC ICM C07K001-00
ICS G06F017-30

AB JP 10287696 A UPAB: 19990113
A database containing the information for amino acid sequence of **protein** with at least biological **function** with added a score on importance of expression of the biological information for each amino acid residue, is new.

USE - The database is useful for determination of unknown biological function of a **protein** or **polypeptide** based on the homology of amino acid sequence, e.g. steric structure of **protein**, and includes retrieval and evaluation of high homologous relationship for the determination of mostly resembling **protein**.

ADVANTAGE - The database allows for correct and rapid retrieval and presumption of **protein** and **polypeptide** having biological **functions**.

Dwg.1/0

FS CPI EPI
FA AB; GI
MC CPI: B04-C01; B04-N03; B11-C08; B12-K04A
EPI: T01-J05B

L123 ANSWER 46 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 1998-437455 [37] WPIX
DNN N1998-340768 DNC C1998-133068

mean), and from this information the **functional** site or sites can be predicted.

Also claimed are:

- (1) an apparatus (e.g. computer systems) for automating this process;
- (2) three heat-resistant DNA synthetase enzymes (sequences given in the specification) obtained by modification of *Pyrococcus furiosus* DNA synthetase at its **functional** sites using the information obtained by this process;
- (3) DNA coding for (2);
- (4) vectors containing (3);
- (5) transformant *Escherichia coli* containing (4) (*E.coli* HMS174(DE3)/pLP320 (FERM P-16712), *E.coli* HMS174(DE3)/pDP501V (FERM BP-6181) and *E.coli* HMS174(DE3)/pDP5C4 (FERM BP-6191)); and
- (6) a preparation of the heat-stable enzymes by culture of the transformants.

USE - The process and the materials are used for the prediction of the **functional** sites of a **protein** from **genomic** or cDNA analysis data and use of this information to produce **proteins** of improved **functionality** by mutation at the **functional** sites so identified.

Dwg.16/24

FS CPI EPI
 FA AB; GI
 MC CPI: B04-E02E; B04-E03; B04-F10A3E; B11-C08E4; B12-K04F;
 D05-A01F; D05-H09; D05-H12B2; D05-H12E; D05-H14A1;
 D05-H17B3
 EPI: S05-C; T01-J01A

L123 ANSWER 47 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 1998-383201 [34] WPIX

DNN D1998-383201 DNC C1493-117120

TI Determining spatial structure of **proteins** from analysis of accessibility of **protein** side chains - then forming and refining model structures from conformational energies, used e.g. in pharmaceutical development.

DC B04 B16 S03 T01

IN NOELTING, B

PA (NOEL-I) NOELTING B

CYC 1

FI DE 19756284 A1 19980716 (199834)* 8p G01N033-68

ADT DE 19756284 A1 DE 1997-19756284 19971211

PFAI DE 1997-19756284 19971211

IC ICM G01N033-68

ICS G01N037-62; G01N036-50; G06F017-50

AB DE 19756284 A UPAB: 19980826

The spatial structure of a **protein** (I) of known primary structure is established by:

- (a) determining the surface accessibility (SA) of **protein** side groups;
- (b) determining conformations that are compatible with SA and physically possible, and
- (c) forming and refining a structural model.

SA is examined by mutagenesis and step (c) involves analysis of conformational energies.

USE - The method is used in molecular biology, medicine, biotechnology, genetics and pharmacy, e.g. to study interactions between enzymes and their substrates or inhibitors, or between **proteins**.

The defined structures can help to understand the effects of biomolecules and to develop pharmaceuticals that modulate the **function** and properties of biomolecules.

ADVANTAGE - The method can be used where usual methods (nuclear magnetic resonance or X-ray diffraction) are not suitable; it is based on the fact that, apart from free energy and partial volume, SA is an important co-ordinate in structure determination. Including SA reduces the number of physically acceptable conformations so much that computer modelling becomes practicable. Good quality structures are produced

quickly with few material requirements.

Dwg. 3/31

ES CPI EPI

FA AB

MC CPI: B04-L01; B04-N04; B11-C08; B12-K04;
D05-H09

EPI: S03-E10; S03-E14H; T01-J15

L123 ANSWER 48 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 1998-11-4934 (24) WPIX

DNN N1998-254159

TI Structural alignment method for comparative analysis of protein structures - using double dynamic programming algorithm.

DC T01

IN TOH, H

PA (PIOM-N) BIOMOLECULAR ENG RES INST; (SEIB-N) SEIBUTSU BUNSHI KOGAKU KENKYUUCHO KP

CYC 7+

PI EP 84-090 A2 19980614 (199829)* EN 40p G06F017-30 <--
P: AL AT BE CH DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO
SE SI

JP 10185925 A 19980714 (199838) 21p G01N033-68

US 6125351 A 20000406 (200001) G06F019-00 <--

ADT EP 84-090 A2 EP 1998-710365 19971219; JP 10185925 A JP 1998-340727
19981110; US 6125351 A US 1997-992176 19971217

PRAI JP 1998-340727 19981220

REP NO-SP.PUB

IC ICM 001M013-00; G06F017-30; G06F019-00
ICC 001M011-00; G06F017-00

AB EP 84-090 A 19980714

The structural alignment method involves using a double dynamic programming algorithm. Distance cut-off approximation is performed, in which a sphere having a predetermined radius and centred on the side chain of a residue of a protein is defined, and residues with side chains centres that are present within the sphere are selected as constituent elements of a structural environment of the residue.

A cut-off approximation is performed for selectively comparing residue pairs obtained by the distance cut-off approximation and having similar local environments.

USE - Comparative analysis of protein structures to obtain information regarding structure, function and evolution of proteins.

ADVANTAGE - Enables structural alignment to be constructed in reduced period of time.

Dwg. 3/31

ES EPI

FA AF; SI

MC EPI: T01-J15H

L123 ANSWER 49 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 1998-11-72130 (24) WPIX

DNN N1998-215014 INC C1998-084971

TI Evaluating structure of proteins by classification of codons in corresponding nucleic acids - used to identify important or mutation-prone regions, functionally similar proteins or coding regions.

DC B04 D16 T04

IN FRESCO, I. R; HALITSKY, D

PA (SUMI-N) CUMULATIVE INQUIRY INC

CYC 7+

PI WO 9606814 A1 19980607 (199824)* EN 56p C07K001-00 <--

PW: AT BE CH DE DK EA ES FI FR GB GH GR IE IT KE LS LU MC MW NL OA PT
SD SE SG EW

W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CZ DE DK EE ES FI GB GE GH
HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW
MX NJ NZ PL PT RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN

YU SW

AU 9350961 A 19980522 (199846) C07K011-00 <--
 ADT WO 9810814 A1 WO 1997-0319673 19971027; AU 9850961 A AU 1998-50961
 19971027

PDT A' 9850961 A Based on WO 9318814

IRAI US 1998-37251 19970203; US 1998-29521 19961026

IC ICM C07K001-00

ICC G06F017-50

AB WO 9318814 A UPAB: 19980911

Structure of a **protein** (I) is evaluated by:

(i) assigning bases in the nucleic acid sequence (II) encoding (I) into triplets, and

(ii) assigning these triplets into classes of a binary choice alphabet of n degrees of freedom, these classes being generated by applying n binary choice parameters (which are **functions** of a message-level properties of (II)) to a triplet to yield second classes.

Also claimed are:

(i) a class-constant table of nearest neighbour (NN) relationships for amino acid (aa) residues that provides, for each class of class-constant NN, a frequency of occurrence for each NN pair in a collection of at least 10 **proteins**;

(ii) a machine-readable medium on which these tables are stored, and
 (iii) a method for evaluating a **protein**:

(i) for resistance to **evolutionary** or mutational change by identifying regions encoded by runs of a single subclass, or

(ii) for presence of critical aa by identifying minority codons in runs of codons of a single class or subcode.

NOTE - The method is used:

(i) to identify **functionally** and structurally important regions of (I);

(ii) to detect coding regions;

(iii) to identify structural and **functional** similarities in **proteins**, and

(iv) to identify mutation-prone regions of viral (II).

Particularly it is used for design of new **proteins** with selected structural or **functional** properties, especially biological activity.

ADVANTAGE - The method can identify structural homologies between **proteins** even where sequence homology is low.

Dwg.0/0

ES CPI EPI

EA AB

MC CPI: B04-E01; B04-N04; B11-C07; B12-K04;
 D05-H09; D05-H18A

EPI: T04-A02B

1123 ANSWER 50 OF 73 WPIX COPYRIGHT 1991 DERWENT INFORMATION LTD

AN 1998-297330 (199846) WPIX

DNC C1998-065404

TI DNA encoding calcium-activated potassium channel - useful in assays to identify compounds which increase or decrease potassium ion flux.

DC B04 D16

IN APelman, J P; Bond, C T; Maylie, J; Silvia, C P

PA (ICAG-N) ICAGEN INC; (UYYOF-N) UNIV OREGON HEALTH SCI; (ICAG-N) ICA-GEN INC

CYC 50

PI WO 9811134 A1 19980314 (199816)* EN 151p C07K014-705 <--

RW: AT BE CH DE DK EA ES FI FR GB GH GR IE IT KE LS LU MC MW NL OA PT
 ND SE SG UW ZW

W: AL AM AT AU AZ BA BB BG BY CA CH CN CU CZ DE DK EE ES FI GB GE
 GH HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN
 MW MK NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ
 VH YU ZW

AU 9742660 A 19980401 (199816) C07K014-705 <--

EP 948542 A1 19991013 (199947) EN C07K014-705 <--

R: AT BE CH DE DK ES FI FR GB GH IE IT LI LU MC NL PT SE

JP 2000514310 W 20001031 (200059) 156p C12N015-09

AU 726159 B 20001102 (200062) C07K014-705 --
 ADT WO 9811139 A1 WO 1997-0316033 19970910; AU 9742660 A AU 1997-42660
 19970910; EP 94-541 A1 EP 1997-0316033 19970910, WO 1997-0316033 19970910;
 JP 2000514310 W WO 1997-0316033 19970910, JP 1993-017828 19970910; AU
 726158 B AU 1997-42660 19970910
 FDT AU 9742660 A Based on WO 9811139; EP 94-542 A1 Based on WO 9811139; JP
 2000514310 W Based on WO 9811139; AU 726159 B Previous Publ. AU 9742660,
 Based on WO 9811139
 PRAT US 1997-45123 19970917; US 1996-26451 19960911; US 1997-46052
 19970907
 IC ICM C07K014-705; C12N013-09
 ICM C07K016-00; C07K016-28; C12N01-15; C12N01-14;
 C12N01-21; C12N01-12; C12N01-09; C12N01-12; C12N01-6;
 C12N01-12; C12N01-01; C12N01-6; C12N01-15; C01N13-02;
 C01N013-03; G06F017-30
 AB WO 9811139 A (PAPB: 19981102
 A new nucleic acid (i) encodes a monomer of a calcium-activated potassium channel (A), where the monomer: (i) has a calculated molecular weight of between 40 and 60 kDa; (ii) has a unit conductance of between 3 and 60 pS when the monomer is in the **functional** polymeric form of a potassium chain and is expressed in a *Xenopus* oocyte; and (iii) specifically binds to antibodies generated against the following sequence, where the C-terminal R is optionally absent: *ATKLELTWAEKRVHNFWMMDTQUTK(R)*
 All sequences are given in the specification. Also claimed are: (1) an isolated nucleic acid encoding at least 11 contiguous amino acids from (A), which has one of 9 sequences; and conservatively modified variants, with the proviso that the contiguous amino acids do not consist of a glutamine repeat amino acid sequence; (2) an isolated nucleic acid of at least 10 nucleotides which specifically hybridises, under stringent conditions, to a nucleic acid encoding an intermediate (A) having a 428 amino acid sequence or (A) which has one of 9 sequences; (3) an isolated (A) having at least 15 contiguous amino acids from one of 9 sequences, their conservatively modified variants, where the variants specifically react, under immunologically reactive conditions with an antibody reactive to one of the 9 **proteins**; (4) an antibody specifically reactive to (A) which has one of 9 sequences; (5) an expression vector comprising a nucleic acid encoding (A) as described above; (6) a host cell transfected with the vector of (5);
 USE The antibody and probe can be used to detect the presence of (A) or (I) in a sample. Host cells expression (A) can be used in assays to identify compounds which increase or decrease the potassium ion flux through (A). The transfected host cell can also be used for the recombinant production of (A). The nucleic acid sequences encoding (A) can also be used for determine mutation sin the SK and IK genes in a computer system. The **proteins** encoded by the SK and IK genes can be used in a computer system for determining their three dimensional structure, which is useful for determining ligands that bind to the **proteins**.
 . (All claimed).
 Dwg.0/0

FS CPI
 FA AB
 MC CPI: B04-E01; B04-F01; B04-G01; D05-H09; D05-H11; D05-H12A;
 D05-H12D1; D05-H12E; D05-H14

L123 ANSWER 51 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 1998-145797 [16] WPIX

DNC C1998-047743

TI Identifying optimal molecular structures by variable basis Monte Carlo method - particularly to determine biologically active conformation(s) of neurotrophin domains involved in receptor binding and subsequent **evolution** of active ligands.

DC B04 T01

IN BLOPELLE, R J; BOSS, G M; SHAMOVSKY, I L; WEAVER, D F
 PA (TOOH) UNIV QUEENS KINGSTON

CYC 7

PI WO 9806048 A2 19980212 (199813)* EN 170p G06F017-50 <--

PW: AT BE CH DE DK EA ES FI FR GB GH GR IE IT KE LC LU MC MW NL OA PT
SD SE SG TG TW

W: AL AM AT AU AZ BA BB BG BR CA CH CN CU CZ DE DK EE ES FI GB GE
GH HU IL IS JP KE FG KP KP LC LF LR LS LT LU LV ML MG MK MN MW
MX NO NC PL PT RU FU SD SE SG SI SE SL TJ TM TR TT UA UG US UZ VN
YU ZW

AU 9736174 A 19980225 (199829) G06F017-50 --

US 6013114 A 20000112 (200017) G06F019-00 --

ADT WO 9806046 A2 WO 1997-CAT39 19970731; AU 9736174 A AU 1997-36174 19970731;
US 6013114 A US 1997-304446 19970731

FDT AU 9736174 A Based on WO 9806046

IRAI GB 1996-16110 19960731

IC ICM G06F017-50; G06F019-00

ICS C07K014-00; G06F017-00

AB WO 9806046 A (PAB: 19980324)

Variable basis Monte Carlo (VMMC) simulated annealing method for identifying an optimal molecular structure comprises (a) providing a Markov chain with an initial basis set of N configuration variables defining a structure; (b) translating, by an amount randomly chosen within a predetermined range, these variables along a basis vector (BV) to produce a new structure; (c) calculating the potential energy (PE) of this structure; (d) deciding, on the basis of a temperature-dependent transition function, whether or not to accept this structure, and if it is accepted using it to replace the existing structure; (e) repeating (b)-(d), each time with a new BV for a specified number of repetitions at a preselected upper temperature; (f) decreasing the temperature according to a preselected cooling schedule; (g) repeating (b)-(a), and (f), each repetition having a new BV for a preselected number of repetitions; (h) storing current molecular structures; (i) repeating (g) and (h) until a predetermined number of structures has been stored; (j) rotating the set of conformational variables according to a distribution of stored structures such that the BV is directed along low energy valleys of a PE hypersurface, so accelerating conformational motions and structural transitions, while erasing structures from (h) and (i); (k) repeating (i) and (j) until, after an appropriate number of repetitions, a preselected low temperature is reached, resulting in an unrefined global minimum structure; (l) refining this structure to identify at least one local minimum on the PE hypersurface. Also new are (1) ligands (I) that bind the TrkB and B receptors containing at least 3 elements able to bind to 3 binding areas in the receptor; (2) ligands (II) that bind the common neurotrophin receptor p75NTR containing at least 2 elements that bind to 2 binding areas in this receptor; (3) designing (I) and (II) by computational evolution.

USE - The method is used to identify the biologically active conformations of **peptide** domains of ligands, particularly neurotrophins (NT), that bind to a receptor. Once such conformations have been identified, small molecules that either inhibit NT-binding or act as agonists of the native ligand can be developed.

ADVANTAGE - The VMMC is more effective than known Monte Carlo methods because (i) translation about the BV is limited and (ii) rotation of the basis set is avoided at high temperature and only introduced at low temperature.

PREFERRED PROCESS - BV is chosen randomly or deterministically within a predetermined range, and the number of repetitions in (g) is at least 2N. The refining step (l), is non-linear minimisation, using a gradient descent, conjugate gradient, quasi-Newtonian (Newton-Raphson), Marquardt, variable metric or Powell minimisation. The molecular system particularly includes at least one **peptide** domain and the basis set includes torsional angles between alpha-carbon atoms of each amino acid and neighbouring C and N atoms, but keeping such angles that describe planar parts of the structure constant. Equations describing the temperature-dependent transition function are presented, and the upper and lower temperatures are 1000 and 120 K. To identify a binding domain, (nearly) optimised structures of domains in a large number of ligand analogues (at least one being able to bind the receptor) are identified, as local PE minima, using the new VMMC method, then these

structures compared to detect structural features common to active ligands. Specifically the method is applied to (a) nerve growth factor/TrkA; (b) brain-derived neurotrophic factor, NT3 or 4/TrkB; (c) NT3/TrkC and (d) the receptor p75NTR. PREFERRED LIGAND - (I) comprises elements with proper spatial occupancy, relative atomic positions, bond type and charge for defining a 3-dimensional configuration able to bind the 3 binding areas. TrkA includes a leucine-rich motif (LFM, amino acids 93-117, with 5 binding areas: (A) Phe111tA (hydrophobic interaction); (B) Phe111tA, Phe111tA and Thr114tA (hydrophobic interaction); (C) Asp109tA and His111tA (ionic); (D) Lys114tA (ionic) and (E) Asn95tA-Ile97tA (multiple parallel beta-strand type hydrogen bonds). Particularly (I) has elements that bind (A), (B) and (C) respectively Trp, Arg and Lys and Asp, optionally also one of D and E, optionally as their isosteres. A similar analysis of binding areas in TrkB and p75NTR is given. ALGORITHM - The specification includes a computer program, in Fortran-77, for implementing VENCO. The algorithm is interfaced with the CHARMM force field so that torsional distortion, van der Waals and electrostatic interactions and hydrogen bonding are accounted for explicitly.

EXAMPLE - A set of 6 **proteins** (3 active and 3 inactive in TrkA binding) was analysed by VENCO to identify low energy conformations of the termini. The algorithm shows that the N (1-11) and C (112'-118') termini of the dimeric (1-118) NGF form a complex that includes a rigid region (1-11 and 112'-118') and a flexible loop (1-8), with the main stabilising factor being ionic contact between Glu11 and Arg11'. Separation of the two regions is caused by repulsion between His4 and His5, Arg10 and Arg11'. This geometrical arrangement is recognised by the receptor. Inactive **proteins** lack the proper geometry of the rigid region.

Fig. 1/13

EP CPI EPI
FL AB
MC CPI: B.4-K11
EPI: T01-T15X

L123 ANSWER 52 OF 73 WPIX COPYRIGHT 1991 DEFENT INFORMATION LTD

AN 1998-065463 (07) WPIX

DNN U1998-051473 DDC C1998-0.1.605

TI Rapid identification of easily synthesised candidate drug compounds - utilising computerised representation of biological macromolecule surface binding site and lists of available reagents to generate virtual libraries of candidates.

DC E04 D16 J04 T01

IN MURRAY, C; YOUNG, S C

PA (EBOT-H) PROTEUS MOLECULAR DESIGN LTD

CYC 13

PI EP 818744 A2 19980114 (1997-7)* EN 24p G06F017-50 <--

R: AD AT BE CH DE DK ES FI FR GB GR IE IT LI LT LU LV MC NL PT FO SE
SI

ADT EP 818744 A2 EP 1997-304412 19970624

PAI GB 1996-16562 19960907; GB 1996-14302 19960718

ICM G06F017-50

ICS C07K001-00; C07K005-08; C12Q009-74

AB EP 818744 A2 EPAB: 19980216

Identifying drug candidate (DC) comprises: (1) obtaining a computerised representation of the 3-dimensional structure of a binding site (BS) on the surface of a biological macromolecule; (2) generating a computerised model (CM) of the **functional** structure of BS which may be used to identify favourable and unfavourable interactions between BS and a DC molecule; (3) identifying a molecular fragment (MF) capable of placement within BS and capable of carrying at least 1 substituent group (SG), where MF is either capable of being synthesised from reagent compounds (RC) accessible in substituted form to introduce SG on synthesis of MF or is present in an RC capable of substitution with SG by reaction with further RC's; (4) generating lists of accessible RC's, such that combination of compounds from each list may be reacted to produce a DC compound containing SG's, to generate a first virtual library (VL) of DC's

especially the theoretical list of compounds producible by reaction of the members of the lists, each member of each list containing a component common to the other members of that list and a component unique within that list; (5) limiting the number of members of each list using a first set of exclusion rules to generate a restricted second VL of DC's, where operation of the rules involves in viewing for each member of each list comparisons for (un)favourable interactions between CM and a structure comprising MF and an SG deriving from the unique component within the list of that member, MF and CM being held in fixed spatial relationship to each other for the comparison; (6) evaluating and ranking by computer the members of the second VL for favourable/unfavourable interactions with CM and thus generating a restricted third VL of DC's ranked as having favourable interactions; (7) optionally selecting at least 1 further MF from the third VL and repeating steps (4)-(6) to generate an alternative third VL; (8) screening the third VL using a second set of exclusion rules to generate a restricted fourth VL of compounds which are candidates for synthesis and experimental evaluation of drug efficacy; (9) synthesising some or all of the compounds of the fourth VL to produce a DC compound library; (10) experimentally evaluating the DC compounds for drug efficacy; (11) analysing the experimental efficacy data for structure-activity relationship information; (12) using the obtained information to select a revised set of lists of accessible EC's, expanded to include, and optionally restricted to exclude, selected EC's present in the lists generated in step (5); (13) repeating steps (6) and (7) to identify further candidate compounds for synthesis and experimental efficacy evaluation; (14) synthesising and evaluating the further compounds; (15) if required repeating steps (11)-(14) at least 1 time, and (16) identifying as lead candidate a compound synthesised and experimentally evaluated as above. Also claimed are: (A) novel active compounds identified by the above process, and (B) a method for manufacturing a drug comprising steps (1)-(16) as above followed by: (a) manufacturing the compound identified in step, and (b) optionally admixing the product with at least 1 carrier or excipient.

ADVANTAGE - A relatively small set of readily synthesisable DC's is rapidly generated, with a high success rate in terms of drug efficacy and hence a high predictive value for directing subsequent iterations. The synthesis is simplified and minimised and the success rate is maximised. The identified DC's are implicitly readily synthesised.

Page 0/3

FS CPI EPI
FA AB
MC CPI: B04-M01; B12-K04; D05-H09; J04-B01
EPI: T01-J15X

LI23 ANSWER 53 OF 73 WPIX COPYRIGHT 2001 DREWENT INFORMATION LTD
AN 1997-374923 [35] WPIX
DNC C1997-120969
TI Polynucleotide(s) and **proteins** derived from *Staphylococcus aureus* - stored on computer readable medium and used in the production of anti-*S.aureus* vaccines.
DC B04 D16
IN BARASH, S C; CHOI, G; DILLON, P J; FANNON, M R; KUNSER, C A; POSEN, C A; CHOI, G H
PA (HUMA-N) HUMAN GENOME SCI INC
CYC 25
PI EP 7-66519 A2 19970706 (199707) * EN C12N015-00
F: AL AT BE CH DE DK ES FI FF GR GR IE IT LI LT LU IN MC NL PT RO SE
SI
CA 2194411 A 19970706 (199707) C12N015-01
JP 09322781 A 19971216 (199803) C12N015-09
ADT EP 7-66519 A1 EP 1997-100117 19970107; CA 2194411 A CA 1997-2194411
19970106; JP 09322781 A JP 1997-20160 19970106
PRAI US 1,66-9561 19960105
REP No-SF.Pub
IC ICM C12N015-00; C12N015-09; C12N015-31
ICS A61K039-085; A61K039-395; C07H021-04; C07K014-31;

C07K016-12; C12N001-19; C12N001-21; C12N005-19; C12Q001-62;
G01N033-569; G06F017-30; G06F019-00

ICA C12P021-02; C12P0_1-08

ICI C12N015-00, C12F011-4

AB EP 736519 A UPAB: 19970828

The following are new: (a) computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM, having recorded on it a nucleotide sequence of the *Staphylococcus aureus* **genome** having one of the 5191 sequences given in the specification, a fragment of these, or a sequence which is at least 95% identical to these sequences, or degenerate variants; and (b) a computer-based system for identifying fragments of the *S.aureus* **genome** of commercial importance comprising: (i) a data storage means comprising the 5191 nucleotide sequences above, a fragment or a sequence which is at least 95% identical to these sequences; (ii) search means for comparing a target sequence to the nucleotide sequence of the data storage means to identify homologous sequences; and (iii) retrieval means for obtaining the homologous sequences.

THE - Homology searches using the new *S.aureus* DNA sequences allows putative **functions** to be assigned so that **protein** -encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are likely to encode antigens have been identified and these **polypeptides** can be used in a vaccine composition against *S.aureus* infection (claimed). The **polypeptides** can also be used in a claimed kit for the immuno-detection of *S.aureus* in a sample. *S.aureus* is implicated in numerous human diseases, including cellulitis, eye-in infections, food poisoning, ecthyma, skin and surgical wound infections, scalded skin syndrome, toxic shock syndrome, etc. The transformed organisms can be used for recombinant production of the **polypeptides**. The new DNA sequences (and their fragments) are useful as primers or probes for isolating homologues of any of the 5191 *S.aureus* sequences (claimed).

DW1.0/2

FS CPI

FA AB

MC CPI: B04-E03; B04-E07F; B04-E08; B04-G01; B04-N03; B04-P0100E; B04-P01E; B14-C11B; D05-H04; D05-H07; D05-H12A; D05-H12D1; D05-H12D5; D05-H12E; D05-H14; D05-H17A5

L123 ANSWER 54 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 1997-244768 (22) WPIX

DNN N1997-201912 DUC C1997-079164

TI Preparing database of molecular fragments by counting all fragments in a molecule - and storing counts in computer memory, useful for analysing structure-activity relationships, especially of drugs and toxins.

DC B04 B04 S03 T01

IN BONIE, F G A; VILLAR, H O

PA (TERR-1) TERRAPIN TECHNOLOGIES INC

CYC 26

PI WO 9714106 A1 19970417 (19972) * EN 48p G06F017-30 <--

EW: AT BE CH DE DK ES FI FR GB GR IE IT LU MC NL PT SE

W: AU CA JP

AU 9673987 A 19970430 (19974) G06F017-30 <--

ADT WO 9714106 A1 WO 1996-US16196 19961010; AU 9673987 A AU 1996-73987 19961010

FDT AU 9673987 A Based on WO 9714106

PRAI US 1995-550647 19951031; US 1995-542642 19951013

REP 1.0n.Ref; EP 496902

IC ICM G06F017-30

IUS G1N031-00; G06F017-50

AB WO 9714106 A UPAB: 19970530

Database of molecular fragments is prepared by:

(a) identifying all sequentially attached fragments within a selected molecule,

(b) counting the occurrences of each unique fragment and

(c) storing information correlating fragment counts with fragment

identity in computer-readable form.

Also claimed are:

- (1) data processing system for creating such databases, and
- (2) computer-readable medium on which the databases are stored.

USE - Comparison of fragment counts between a molecule and a reference molecule of known activity can be used to predict which compound will have this particular activity. The method is especially applied to libraries of drugs (e.g. central nervous system drugs), toxins or randomly chosen compounds.

ADVANTAGE - The databases provide a complete and systematic classification of **function** activity based on specific topological characteristic of fragments of small molecules, and is suitable for construction of combinatorial libraries covering the whole of chemical space or focused on part of it for precise selection of active molecules.

Dwg.13a/13

FS CPI EPI

FA AB; GI

MC CPI: S04-E01; B12-K04A; C04-B
EPI: S03-E14A1; S03-E14H; S03-E14X; T01-J05B2A; T01-J05B3;
T01-J10C4

L123 ANSWER TO CF 73 WPIX COPYRIGHT 1991 DOWWENT INFORMATION LTD

AN 1396-1300839 (36) WPIX

DNC 01396-1305673

TI Predicting tendency to form amphiphilic alpha and beta structures - using a novel algorithm to calculate values for subsequent graphical analysis to predict **protein** structure.

DC B14 T01

IN CHEUNG, M; ODEGLEDY, F; FINCHBORG, J; ICEROWITCH, P; LI, J
PA (UTCG) UNIV COLUMBIA NEW YORK

CYC 31

PI WO 9618957 A1 19960620 (199606) EN 16p G06F17-10 <--
FW: AT BE CH DE DK ES FR GB GR IE IT LU MC NL PT SE

W: AU CA JP

AC 9644684 A 19960703 (199641) G06F17-10 <--
US 5940307 A 19990817 (199908) G06F17-11 <--

ADT WO 9618957 A1 WO 1996-US16126 19951213; AU 9644684 A AU 1996-44684
19951213; US 5940307 A US 1994-355-44 19941214

FDT AC 9644684 A Based on WO 9618957

PRAI US 1994-355844 19941214

REP 6.Jnl.Ref; US 4853471; US 4939666; US 5265030

IC ICM G06F017-10; G06F017-11

ICS G06F017-50; G06F019-00

AB WO 9618957 A WPAR: 1396071

Predicting the tendency of a **protein** to form either: (a) an amphiphilic alpha-structure or (b) an amphiphilic beta-structure comprises: (i) calculating a series of values for (a) U-alpha-x using equation (1) or (b) U-beta-x using equation (2), for a series of portions of the **protein**, each portion having a span of x residues, where the series of portions spans the **protein** and x is an integer:

Equation (1) is U-alpha-x = Hx + mu-alpha-x - [pt] and Equation (2) is U-beta-x = Hx + mu-beta-x - [pt]. In equations (1) and (2), Hx is the average hydrophobicity for a span of x residues using the Kyte-Doolittle scale, mu-alpha-x is the hydrophobic moment (span x) for alpha-structures, the angle between one residue and the successive residue being 100 deg., mu-beta-x is the hydrophobic moment (span x) for beta-structures, the angle between one residue and the successive residue being 160 deg., and [pt] is the position dependent turn propensity, and (ii) depicting the values for U-alpha-x or U-beta-x graphically to form a series of peaks, where (a) peaks wide enough to correspond to a segment of the amino acid sequence long enough to span the membrane as an alpha-helix are predicted to be alpha-structures and (b) peaks that are too narrow to correspond to a segment of the amino acid sequence long enough to span the membrane as an alpha-helix but which are wide enough to correspond to a segment of the amino acid sequence with a length between 6 and 14 amino acid residues are

predicted to be beta-structures.

USE - The methods are used partic. for predicting the structure of membrane **proteins** such as glucose transporter **proteins** (GLUTS). They can be used to discern the **function** of **proteins**. They can also be used for the rational design or identification of cpds. which interact with the **proteins** or to engineer **proteins** having particular structures.

Dwg.07/

FS CPI EPI

FA AB

MC CPI: B04-N04; B12-K04
EPI: T01-J00A; T01-S

L123 ANSWER 36 OF 73 WPIX COPYRIGHT 1991 DERMONT INFORMATION LTD

AN 1995-086705 (C1, WPIX

CR 1997-525719 (48)

DNN N1995-005454 DMC C1945-900356

TI Analysis and modification of **protein** structure - by identification of favoured and suppressed patterns of hydrophobic amino acid residues.

DC B04 D16 T01

IN HUMPHREYS, P E

PA (ANTI-H) ANTIGEN EXPRESS INC

CYC 19

PI WO 9426783 AI 19941124 (1995)1* EN 18p C07K013-00 <--
FW: AT BE CH DE DK ES FR GB CR IE IT LU MC NL PT SE
W: CA JP

AST WO 9426783 A1 WO 1994-US5552 19940518

FFAI US 1993-63908 19930519

REP 0!Jnl.Lef

IC ICM C07K013-00

ICS G06F015-00

AB WO 9426783 A UPAB: 19971209

(A) Identification of favoured and suppressed patterns of hydrophobic residues from the group Leu, Ile, Val, Phe and Met indicated F and other residues indicated S, comprises: (a) providing amino acid (AA) sequences from naturally occurring **proteins**; (b) applying templates of 3 to 9 positions composed of all combinations of F and S in each position of the amino acid sequences of (a); (c) scoring the frequency with which each template is found; (d) determining the standard deviation (SD) by which a template occurs from the expected frequency for an empirical distribution of F and S positions based upon the frequencies of the AAs in the naturally occurring **proteins** of (a); and (e) ranking the patterns on a scale of preferred and suppressed on the base of those SDs between observed and expected frequencies. Also claimed are: (B) alteration of the structure of a predetermined AA copolymer, comprising: (a) providing the sequence of AA residues in an AA copolymer; (b) providing the identity of favoured and suppressed patterns of hydrophobic aa residues as determined by (a)-(e) as in (A); (c) identifying patterns of hydrophobicity in the provided AA copolymer on the basis of (b); (d) comparing the concordance of the observed AA residues to patterns which have been determined to be favoured or suppressed; and (e) altering the identity of the observed aa residues to effect a structural change; (C) identification of the extent and density of the pool of hydrophobic side chains in the core of a **protein** into which hydrophobic side chains of an alpha-helix are immersed, comprising: (a) providing X-ray crystallographic coordinates of naturally occurring **proteins**; (b) applying the strip-of-helix template (X---X--X---X---X---, joined in a circle) to each of the alpha-helices in the **proteins** of (a) to maximise the mean hydrophobicity of residues in X positions; (c) extending the template pattern defined by the application in step (b) to adjacent non-helical regions; (d) determining the template-specified positions from the gp. of (i) all X positions in the helix, (ii) the N-terminal X position, (iii) the C-terminal X position, (iv) the X positions in the helix between the N-terminal X position and the C-terminal X position, (v) the X position with the smallest residue, (vi) the X position with the

least hydrophobic residue, (vii) the first X position beyond the helix at the N-terminus, (viii) the first X position beyond the helix at the C-terminus, (ix) all non-X positions in the helix and (x) the X position contg. one of two sequential residues in the primary sequence from the gp. Lys, Ile, Val, Phe and Met; (e) determining the identity of all AAs as a function of the distance of their C- alpha, from the distal non-hydrogen atom of the side chain of AAs in the template-defined positions and (f) determining the mean density, cumulative hydrophobicity and shell hydrophobicity of each residue identified in step (e) around each template-defined position; (ii) altering the structure of an alpha-helix and/or altering the structure of the appositional hydrophobic pool to that helix, in a predetermined AA copolymer, comprising: (a) providing a sequence of AAs from a predetermined AA copolymer; (b) step (b)-(f, as in (C)); (c) comparing the concordance of the observed AA residues in the positions of the strip-of-helix template-lines position and in the appositional pool of residues to patterns which have been determined to most favour or most suppress helix extension; and (d) altering the identity of the observed AA residues to effect a structural change; (E) predicting open reading frames (ORFs) in DNA sequences, comprising: (a) providing a DNA sequence to be analysed for the presence of ORFs; (b) determining the DNA codon pattern corresponding to SFFSS and SFSFS, the respective DNA codes being -nonT--T--T--nonT--nonT- and -nonT--T--nonT--T--nonT-, where T is a deoxyribonucleotide with the base thymine and nonT is a deoxyribonucleotide with a base selected from adenine, cytosine, guanine and thymine; (c) scoring the occurrence of each pattern within the DNA sequence; (d) selecting polynucleotide segments which are bounded by DNA nucleotide patterns corresponding to this SFFSS aa pattern and which hold the SFFSS pattern more than one SD from the empirical frequency of that pattern in a standard set of DNA sequences and which exclude the SFSFS pattern more than one SD from the empirical frequency of that pattern in a standard set of DNA sequences; and (e) identifying segments contg. the individual or merged positive selections as the predicted ORFs; (F) predicting of alpha-helices, comprising: (a) providing an AA sequence to be analysed for the presence of a helix; (b) scoring the segments in which the SFFSS pattern is found; (c) extending the scored segment toward the N-terminus and/or the C-terminus to include additional residues which fall within favoured composite patterns which are identified with this invention to include the highly favoured pattern; (d) identifying individual or overlapping segments identified in this manner as the predicted alpha-helices.

The methods can be used for the analysis and modification of **protein** structure. They can be used to design **polypeptides** *de novo* or to modify natural prods. to provide the most efficiently folded and stable forms of AA copolymers. The technique can be applied to drug design.

Dwg.0/3

FS CPI EPI

FA AB; GI

MC CPI: B04-C01; B04-E01; B04-N04; B10-B02; B12-K04;
D05-H12B1; D05-H12D1

EPI: T01-J03

L123 ANSWER 57 OF 70 WPIX COPYRIGHT 2001 REFVENT INFORMATION LTD

AN 1994-120330 [15] WPIX

DNC C1994-055596

TI Method of resurfacing of rodent antibodies to produce humanised antibody forms - for producing non human antibodies with improved therapeutic efficiency by presenting human surface on V-region.

DC B04 D16

IN GUILD, B C; PEDERSEN, J T; REEC, A R; ROGUSKA, M A; SEAFLE, S M J
(PEDE-I) PEDERSEN J T; (IMMU-N) IMMUNOGEN INC

CYC 16

PI EP 592106 A1 19940413 (199413)* EN 23p C12N15-13

E: BE CH DE DK ES FR GB IE IT LI LU NL SE

CA 2105644 A 19940510 (199421) C07K15-28 <--

JP 07067638 A 19950314 (199519) 134p C12P021-08

US 5639641 A 19970617 (1997/30) 158p C12N015-00
 ADT EP 52106 A1 EP 1993-307051 1430907; CA 2105644 A CA 1993-2105644
 19930907; JP 07061088 A JP 1993-114702 19930908; US 5639641 A US
 1992-942145 19920909

PRAI US 1992-942245 19920909

REP 91011.ref; EP 519596; WO 9109067

IC ICM C07K015-28; C12N015-00; C12N015-13; C12P011-06

ICS A61K039-395; C07K015-00; C07K016-00;
 C07K016-46; C12N015-01; C01N023-20; C01N037-68;
 G06F015-42

ICA G1M033-531

AB B1 B91106 A UPAB: 19970723

Humanising a rodent antibody Ab, or fragment by resurfacing comprises:
 (a) determin. of the conformati. nai structure of the variable region of the
 rodent Ab or fragment by constructing a 3D model of the rodent Ab variable
 region; (b) generating sequence **alignments** from relative
 accessibility distributions from X-ray crystallographic structures of the
 rodent Ab variable region heavy and light chains to give a set of heavy
 and light chain framework positions; (c) defining a set of heavy and light
 chain surface exposed amino acid residues using the set of framework
 positions generated in (b); (d) identifying from human antibody amino acid
 residues that is most closely identical to the set of residues defined in
 (c); (e) substituting the set of heavy and light chain surface exposed
 amino acid residues defined in (c) with the set of residues identified in
 (d); (f) constructing a 3D model of the rodent antibody variable region
 resulting from the substitution of (e); (g) identifying any amino acid
 residues from the set identified in (d), that are within 5 Angstrom of any
 atom of any residue of the complementarity determin. regions (cDRs) of the
 rodent Ab to be humanised; and (h) changing any residue identified in (g)
 from the human to the original rodent amino acid residue. Step (a) need
 not be conducted first, but must be conducted prior to step (g).

USE - The humanised antibodies permit the prodn. of
functional altered Abs, which retain the binding parameters of the
 original non-human antibody, with improved therapeutic efficiency in
 patients due to the presentation of a human surface on the variable
 region.

Dwq.0/10

FS CPI

FA AB

MC CPI: B04-G01; B11-C08; B12-K04; D05-H11

ABEQ US 5639641 A UPAB: 19970723

A method for producing a humanized rodent antibody or fragment thereof by
 resurfacing, said method consisting essentially of:

(a) generating position **alignments** from relative
 accessibility distributions from x-ray crystallographic structures of a
 pool of rodent and human antibody heavy and light chain variable regions
 to give a set of heavy and light chain variable region framework surface
 exposed positions wherein the **alignment** positions for all rodent
 and human variable regions are at least about 98% identical;

(b) defining for a rodent antibody or fragment thereof a set of heavy
 and light chain variable region framework surface exposed amino acid
 residues using said set of heavy and light chain variable region framework
 surface exposed positions generated in said step (a);

(c) identifying from human antibody amino acid sequences a set of
 heavy and light chain variable region framework surface exposed amino acid
 residues that is most closely identical to said set of rodent surface
 exposed amino acid residues defined in said step (b);

(d) substituting, in the variable region framework amino acid
 sequence of said rodent antibody or fragment thereof, said set of heavy
 and light chain variable region framework surface exposed amino acid
 residues defined in said step (b) with said set of heavy and light chain
 variable region framework surface exposed amino acid residues identified
 in said step (c);

(e) constructing three-dimensional models of said variable region of
 said rodent antibody or fragment thereof and of said variable region of

said rodent antibody or fragment thereof resulting from the substituting specified in said step (d);

(f) comparing said three-dimensional models constructed in said step (e), and identifying any amino acid residues from said sets identified in said steps (b) or (c), that are within 5 Angstroms of any atom of any residue of the complementarity determining regions of said rodent antibody or fragment thereof;

(g) changing any residues identified in said step (f) from the human to the original rodent amino acid residue to thereby define a humanizing set of surface exposed amino acid residues;

(h) replacing the set of rodent antibody variable region framework surface exposed amino acid residues defined in said step (b) with the humanizing set of variable region framework surface exposed amino acid residues defined in said step (g); and

(i) producing said humanized rodent antibody or fragment thereof having binding specificity.

Dwg.6/10

L123 ANSWER 53 OF 73 WPIX COPYRIGHT 2-61 DEGWENT INFORMATION LTD
 AN 1994-093196 [10] WPIX
 CR 1993-018139 [92]
 DNN N1994-064357 EPO C12M15-160
 TI Preparation of improved humanised antibodies - by comparison of consensus and import complementarity determining regions and framework region sequences, e.g. to humanise murine.
 DC B04 D16 T01
 IN CARTEF, P J; PFESTA, L G
 PA (CETH) GENENTECH INC
 CYC 26
 PI WO 9404679 A1 19930903 (1994-09-01) * 12sp C12M15-16
 PW: AT BE CH DE DK ES FR GB GR IE IT LU MC NL PT SE
 W: AU CA JP
 AU 9350631 A 19930815 (1994-09-01) C12M15-16
 US 5821337 A 19931013 (1994046) C07K016-00 <--
 US 6054237 A 29060425 (2906017) A61K039-395
 ALT WO 9404679 A1 WO 1993-US7551 19930820; AU 9350631 A AU 1993-10831
 19930813; US 5821337 A CIP of US 1991-715272 19910614, CIP of WO
 1992-US5126 19920615, US 1993-334373 19930621; US 6054237 A CIP of US
 1991-715272 19910614, CIP of WO 1992-US5126 19920615, Cont of US
 1992-334373 19920621, US 1993-437642 19930609
 FIT AU 9350631 A Based on WO 9404679; US 6054237 A Cont of US 5821337
 PPAI US 1992-334373 19920621; US 1991-715272 19910614; WO 1992-US5126
 19920615; US 1993-437642 19930609
 REP 3.Jnl.Ref; WO 9007561; WO 9102661
 IC ICM A61K039-395; C07K016-00; C12M015-16
 ICS C07K013-00; C07K016-30; C12M005-10; C12P021-08;
 G06F015-00
 AB WO 9404679 A UPAR: 200000606
 A method for making a humanised antibody (Ab) composed of an amino acid sequence of a non-human, import Ab, and a human Ab, comprises the steps of: (a) obtaining the AA sequence of at least a portion of an import variable domain and a consensus human variable domain; (b) identifying the complementary determining region (CDR) AA sequences in both variable sequences; (c) substituting an import CDR sequence for the corresponding human CDR sequence; (d) aligning the AA sequences of a framework region (FR) of the import Ab and the FR of the consensus Ab; (e) identifying import Ab FR residues in the aligned FR sequences that are **non-homologous** to the corresponding consensus Ab residues; (f) determining whether the **non-homologous** import AA residue is reasonably expected to have one of the following effects; (1) non-covalently binds antigen directly; (2) interacts with a CDR; or (3) participates in the VL-VH interface; and (g) for any **non-homologous** import Ab AA residue which is reasonably expected to have at least one of these effects, substituting that residue for the corresponding AA residue in the consensus Ab FR sequence.

A humanised version of murine monoclonal antibody MuMAb4D5 (ATCC CRL

104G) is provided which retains the binding specificity of the parent but is non-immunogenic in humans. The parent Ab is also directed to the extracellular domain of p185(HER2) and has potential for treatment and imaging of e.g. human breast and ovarian cancers. (The single-step humanisation is shown in the fig.)

USE/ADVANTAGE - The method provides an efficient means of humanising antibodies, so as to retain or to improve the affinity of the non-human donor antibody for a given antigen. The antibodies are less antigenic in humans than non-human antibodies.

Dwg.1 7

FS CPI EPI
FA AB; GI
MC CPI: B04-C01; B04-C01; B04-N04B; B12-K04A1; D05-H09; D05-H11A2;
D05-H17A1; D05-H18
EPI: T1-J06A

L123 ANSWER 60 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 1994-067529 [09] WPIX

DNN N1994-051467 DNC C1994-030042

TI Database classification method based on sequence comparisons - by dividing DNA or **protein** sequences into structural words taking into account structural and **functional** relationships.

DC B04 T01

IN WITTING, B

FA (SOFT-1) SOFT GENE KOMBIKUNGS & VERTRIEBSSYSTEM

CYC 8

PI EP 1994-067529 A1 1994-0316 (1994) * DE DEP G06F015-401 <--
P: AT CH DE FR GE IT LI SE

ADT EP 1994-067529 A1 EP 1992-250210 1992-03-12

FRA1 EP 1994-067529 1994-03-12

REP .1.ber

IC ICM G06F015-401

AB EP 1994-067529 A1 EP 1994-0316

Method comprises cleaving the sequences into structural words which include analysable information about structural and **functional** relationships.

The structural words pref. are (a) nucleotide sequences (NucIndex), pref. taking into account the degenerative nature of the genetic code, the DNA recognition sequences for binding **proteins** and allowing mismatches (CN) or (b) aminoacid sequences (AA, IdenIndex and ChemIndex), pref. taking into account their chemical and steric properties for grouping them.

NucIndex is based on the considerations that (i) 10-12 nucleotides form one helical repeat, (ii) the degenerative code occurs mostly in the third, sometimes in the first, position of DNA triplets, (iii) that the binding sites for regulatory **proteins** usually are segments of 5-6 nucleotides, i.e. half a helical repeat (i). Therefore mismatches are allowed on the other half of helical repeats, (iv) the significance of the number of matches (SigA) depends on the sequence length (SigS); and ChemIndex is based on the considerations that (a) **proteins** are either helices (helical repeat of 3-6 AA) or beta-sheet and turn (at least 4 AA), (b) some AA lead to the same structure and similar characteristics. Additionally, frequently occurring structural words are of a lesser significance than words which occur less frequently.

USE/ADVANTAGE - The method is used for database classification of DNA and **protein** sequences. Compared to prior art methods, the new method is faster and allows better detection of sequence similarities.

Dwg.0/0

Dwg.1/0

FS CPI EPI

FA AB

MC CPI: B04-E01; B11-C08; B12-K04

L123 ANSWER 60 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 1993-308503 [39] WPIX

DNN N1993-137572 DNC C1993-136720

TI Gene information testing device to evaluate similarity between aminoacid sequence and reference amino-acid - comprises unit to detect number of longest common letter between the sequences and calculation unit to find ratio of longest common letter detected.

DC B04-B04-S03 T01

PA (FUIT; FUIT.NU LTD

CYC 1

PI JP 0511932 A 19930731 (199339)* 17p C12M001-00

ADT JP 0511932 A JP 1992-11012 19920206

PEFAI JP 1992-1101 19920206

IC ICM C12M001-00

IPC G06F015-40; G06F015-42

AB JP 0511932 A UPAB: 19931113

Gene information testing device comprises a detection unit (10) to detect the number of the longest common letters between the amino acid sequence to be tested and the reference amino acid sequence each expressed by letters, and a calculation unit (11) to calculate the ratio of the number of the longest common letters detected by the detection unit (10) to the number of letters of the amino acid sequence to be tested or the reference amino acid sequence.

USE/ADVANTAGE - Used to evaluate the similarity between an amino acid sequence to be tested and a reference amino acid sequence. It is necessary for the development of medicines, etc. From the data information such as structure, function, etc. of **protein** can be found. In an example, where the amino acid sequence to be tested is expressed by letters "ABCBDAB" and the reference amino acid sequence is "BDCABA". The detection unit detects the number of the longest common letters which is "4"; and the calculation unit calculates the ratio $4/7$ (= 4 - 7) to the number of letters of the amino acid sequence to be tested and $4/7$ (= 4 - 6) to the number of letters of the reference amino acid sequence. Thus, the similarity can be evaluated according to a simple processing mechanism.

Dwg.1a,b/1

PS CPI EPI

FA AB; GI

MC CPI: B04-B04A1; B11-C05; B12-K04A; D05-H09;
D05-H12; J04-B01

EPI: S03-E14H9; S05-C09; T01-J06A

L123 ANSWER 61 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 1993-267046 [34] WPIX

ENN N1993-204684 DNC C1993-116919

TI Analysis of chromosome shape to determine main axis of symmetry - by forming curved main axis from mid points of 2 peak positions of density in image.

DC B04-B04-S03 T01

PA (KAGG) MAT FES INST METALS; (NIKR) NIKON CORP

CYC 1

PI JP 05180832 A 19930723 (199304)* 11p G01N033-50

ADT JP 05180832 A JP 1991-190982 19910705

PEFAI JP 1991-190982 19910705

IC ICM C01N033-50

IPC A61B010-00; G06F015-62; G06F015-70

AB JP 05180832 A UPAB: 19931119

In the analysis a curved main axis is obtained by connecting the middle points of 2 peak points of density value of picture elements along each of lines perpendicular to a linear main axis of an image of chromosome.

USE/ADVANTAGE - Used to determine the main axis of a chromosome when identifying chromosome. When identifying chromosome, an axis of symmetry (main axis) that separate symmetrically a chromosome into 2 blocks is determined, and, then, various analyses (e.g., detection of centromere) are made. Therefore, it is very important to determine properly the main axis in the identification of chromosomes.

In an example, a linear main axis is obtained on an image of chromosome by the conventional technique; a density **profile** is taken along a line (x) perpendicular to the linear main axis; in the graph

indicating the density **profile**, 2 peak values (p1,p2) of density are selected; the middle point (q) of the 3 point (p1,p2) on the x-axis is obtained; the middle point (j) is plotted at the corresponding point on the image of chromosome; in the same procedure as mentioned above, the lines (x) perpendicular to the main axis (a1) are taken at different positions, the middle points (q) are obtained, and they are plotted at the corresponding positions on the image of chromosome; and, by connecting the middle points (q) obtained, a curved main axis (a2) can be obtained. Thus, a curved main axis can be properly obtained even for a chromosome having 2 arms on both sides of the centromere (2).

FS CPI EPI GMPI

FA AB

MC CPI: B04-B04A1; B11-H04; B12-H04
EPI: S04-S14H04; T01-T10E2

L123 ANSWER 61 OF "3" WPIX COPYRIGHT 2001 DEPARTMENT INFORMATION LTD

AN 1993-1992031 [4] WPIX

DNC C1993-087384

TI Selecting oligo-nucleotide probe for identifying allele of polymorphic gene - by choosing the min. number of probes necessary for unequivocal discrimination from mutant forms of consensus sequence.

DC P04 S16

IN BOUQUELET, L; COHEN, P; COHEN, N; DAUSET, C
(REHU) BERTIN & CIE

CYC 21

PI WO 9311262 A1 19930616 (1993.06) EN 4sp C12Q0001-66
FW: AT BE CH DE DK ES FR GR IE IT LU NL PT SE
W: AU CA DE FR INEP 0493482 A1 19930630 (1993.06) FR 26p C12Q0001-66
F: AT BE CH DE DK ES FR GR IE IT LU NL SE

FR 1634738 A1 19930611 (1993.06) 1sp C12Q0001-66

AP 9333550 A 19930628 (1993.06) C12Q0001-66

JP 07501449 W 19930616 (1993.06) C12Q0001-66

ADT WO 9311262 A1 WO 1992-FF1141 19921203; EP 049338 A1 EP 1992-493387
19921203; FR 2664668 A1 EP 1992-14936 19921204; AU 9313550 A AU 1993-33550
19921203; JP 07501449 W WO 1992-FF1141 19921203, JP 1993-503904 19921203

FDT AU 9333550 A Based on WO 9311262; JP 07501449 W Based on WO 9311262

PFAI FR 1991-14996 19911204

REP 3.Jnl.Ref; EP 103909; EP 237362; EP 412893; WO 8904875; WO 8911547; WO
9208117; WO 9210569; WO 9211189; 5.Jnl.Ref

IC ICM C12Q0001-68

ICS C07H021-00; C12N015-11; G06F015-20

AE WO 9311262 A UPAE: 19931116

Process for selecting, from a set of allelic sequences of a polymorphic gene (FG), at least one mutation screen able to specify at least one oligonucleotide probe for distinguishing between all alleles, comprises: (1) selecting all or part of a known consensus sequence (CS) of FG; (2) producing a matrix of mutations with sequences corresponding to known alleles; (3) identifying indiscernible sequences by pairwise comparison (i.e. alleles having the same mutation **profile** in CS) and excluding one member of each pair; (4) identifying and numbering obligatory mutations and 'marker' mutations (i.e. those which are necessary and sufficient to distinguish 2 otherwise identical alleles; set O of obligatory mutations), and (5) producing at least one minimal mutation screen comprising at least these obligatory mutations.

To identify alleles, the screen selected in step (1) which is most suitable for prodn. of probes to differentiate between the alleles is chosen and these probes used for hybridisation typing of the alleles.

USE/ADVANTAGE - This method provides rapid and reliable identification of alleles without requiring a large no. of probes. It can identify homozygotic doublets and differentiate them from heterozygotic doublets.

Dwg.0/0

FS CPI

FA AB

MC CPI: B04-B04A1; B12-H04; D05-H11; D05-H12

ABEQ EP 549338 A UPAR: 19931116

Selecting, from a set of allelic sequences of a polymorphic gene (PG), at least one mutation screen able to specify at least one oligonucleotide probe for distinguishing between all alleles, comprises (1) selecting all or part of a known consensus sequence (CS) of PG; (2) producing a matrix of mutations with sequences corresp. to known alleles; (3) identifying indiscernible sequences by pairwise comparison (i.e. alleles having the same mutation **profile** in CS) and excluding one member of each pair; (4) identifying and numbering obligatory mutations and 'marker' mutations (i.e. those which are necessary and sufficient to distinguish 2 otherwise identical alleles; set \emptyset of obligatory mutations), and (5) producing at least one minimal mutation screen comprising at least these obligatory mutations. To identify alleles, the screen selected in step (5) which is most suitable for num. of probes to differentiate between the alleles is chosen and these probes used for hybridization typing of the alleles.

USE/ADVANTAGE - This method provides rapid and reliable identification of alleles without requiring a large number of probes. It can identify homozygotic doublets and differentiate them from heterozygotic doublets.

Ewg.0/10

ABEQ FR 1684688 A UPAR: 19931122

Process comprises: (1) selecting all or part of a known consensus sequence (CS) of PG; (2) producing a matrix of mutations with sequences corresp. to known alleles; (3) identifying indiscernible sequences by pairwise comparison (i.e. alleles having the same mutation **profile** in CS) and excluding one member of each pair; (4) identifying and numbering obligatory mutations and 'marker' mutations (i.e. those which are necessary and sufficient to distinguish 2 otherwise identical alleles; set \emptyset of obligatory mutations), and (5) producing at least one minimal mutation screen comprising at least these obligatory mutations.

USE/ADVANTAGE - This method provides rapid and reliable identification of alleles without requiring a large no. of probes. It can identify homozygotic doublets and differentiate them from heterozygotic doublets.

Ewg.0/10

L123 ANSWER 63 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 1993-027210 [03] WPIX

DNN N1993-020821

TI Fuzzy interference appts. e.g. for tunnel trenching by shield machine - **multiplies** oblirescence coefft. by membership **function** of conclusion portion of each operating linguistic control rule, with coefft. determined by fuzzy interference to vary weight of importance of given rule.

DC T01 T06

IN TOKIBE, T

PA (MEID) MEIDENSHA CORP; (MEID) MEIDENSHA KK

CYC 2

PI US 5173801 A 19921229 (199303)* Sp G05B013-02
KR 9705457 B1 19970416 (199909) G06F009-44

ALT US 5173801 A US 1990-581770 19900913; KR 9705457 B1 FF 1990-14455 19900913

PFAI JP 1989-230514 19890914

IC ICM G05B013-03; G06F009-44

AB US 5173801 A UPAR: 19930924

The appts. includes an operating variable fuzzy inference portion (10) which receives input values, i.e., a deviated distances X1 and X2 in horizontal and vertical directions between a planned trenching line and an actual direction of advance of the shield machine. A **function** infers a target value of a stroke quantity of each shield machine jack through fuzzy inference. The portion comprises a linguistic fuzzy control rule storage portion (operating rule storage portion (1), a membership **function** storage portion (2), an oblirescence coefft. storage portion (3) and a fuzzy inference mechanism (4).

The operating rule storage contains a number e.g. n linguistic control rules, which receive X1 and X2 as input values and provide the

stroke quantities y_1 , Y_2 , ---, respectively, for inferred values (concluded values) supplied to jact portion of the shield machine.

ADVANTAGE - Accounts for oblivescence of knowledge base in sowing protein.

1/4

FS EPI

FA AB; GI

MC EPI: T01-J163; T06-A05A1

L123 ANSWER 64 OF 73 WPIX COPYRIGHT 2001 PERWENT INFORMATION LTD

AN 1993-018139 (07) WPIX

CH 1993-018139 (10)

CNN N1993-1-8139 DMC 013-8-008-26

TI Humanisation of antibodies - by molecular modelling of the variable domains and alteration by gene conversion mutagenesis.

DC B04 Div. T01

IN CAFIER, P. J.; PRESTA, L. G.

PA (GETH) GENENTECH INC

CYC 12

PI WO 9222653 A1 19921223 (199301) * EN 166p C12N015-1:

FW: AT BE CH DE DK ES FR GR GF IT LU MC NL SE

W: AU CA JP US

AU 9222653 A1 19921223 (199301) C12N015-1

BE 675916 A1 1993-012604 (1993114) FN C12N015-1

F: AT BE CH DE DK ES FR GR GF IT LU MC NL SE

JP 06506267 W 1993-012604 (1993114) 4ip C12P021-0

AU 675916 E 1993-012604 (1993114) C12N015-1

EP 0404668 A1 1993-0904 (199304) EN C12N015-1

F: AT BE CH DE DK ES FR GR GF IT LU MC NL SE

ADT WO 9222653 A1 WO 1992-012604 19920615; AU 9222653 A AU 1992-012604

19920615; EP 040468 A1 EP 1992-012604 19920615, WO 1992-012604 19920615;

JP 06506267 W WO 1992-012604 19920615, JP 1993-011103 19920615; AU 675916

E AU 1992-012604 19920615; EP 040468 A1 Div ex EP 1992-012604 19920615, EP

1993-012604 19920615

FDT AU 9222653 A Based on WO 9222653; EP 590058 A1 Based on WO 9222653; JP

06506267 W Based on WO 9222653; AU 675916 B Previous Publ. AU 9222653,

Based on WO 9222653; EP 040468 A1 Div ex EP 590058

PRAI US 1991-715272 19910614

REP P.Jnl.Ref; WO 9207361

IC ICM C12N015-18; C12P021-06

ICS C07K013-00; C07K016-28; C07K016-46;

C12N005-10; C12N015-02; G06F015-00

AB WO 9222653 A UPAB: 19930126

A method for making a humanised antibody comprising an amino acid sequence of a non-human, import antibody and a human antibody is claimed comprising (a) obtaining the amino acid sequences of at least a portion of an import variable region and of a consensus human variable domain, (b) identifying Complementary Determining Region (CDR) amino acid sequences in the import and the human variable domain sequences, (c) substituting an import CDR amino acid sequence for the corresponding human CDR amino acid sequence, (d) aligning the amino acid sequences of a Framework Region (FR) of the import antibody and the corresponding FR of the consensus antibody, (e) identifying import antibody FR residues in the aligned FR sequences that are **non-homologous** to the corresponding consensus

antibody residues, (f) determining if the **non-homologous** import amino acid residue is reasonably expected to have at least one of the following effects: (i) non-covalently binds antigen directly, (ii)

interacts with a CDR or (iii) participates in the V_H-V_H interface and (g)

for any **non-homologous** import antibody amino acid residue which is reasonably expected to have at least one of these

effects, substituting that residue for the corresponding amino acid

residue in the consensus antibody FR sequence.

Also claimed are: (A) a method comprising (a) providing at least a

portion of an import, non-human antibody variable domain amino acid

sequence having a CDR and a FR, (b) obtaining the amino acid sequence of

at least a portion of a consensus human antibody variable domain having a

CDR and a FF, (c) substituting the non-human CDR for the human CDR in the consensus human antibody variable domain and (d) substituting an amino acid residue for the consensus amino acid residue at at least one of the following sites: 4L, 35L, 36L, 38L, 43L, 44L, 46L, 58L, 62L, 63L, 64L, 65L, 66L, 67L, 68L, 69L, 70L, 71L, 73L, 85L, 87L, 98L, 2H, 4H, 24H, 36H, 37H, 39H, 41H, 45H, 48H, 50H, 57H, 63H, 69H, 70H, 72H, 74H, 75H, 76H, 78H, 91H, 93H and 104H; (E) a humanised antibody variable domain having a non-human CDR incorporated into a human antibody variable domain, where the improvement comprises substituting an amino acid residue for the human residue at a site selected from those defined in (A); (F) a **polypeptide** comprising the amino acid sequence (I) or (II); (G) a method for engineering a humanised antibody comprising introducing amino acid residues from an import antibody variable domain into an amino acid sequence representing a consensus of mammalian antibody variable domain sequences; (H) a computer comprising the sequence data of the amino acid sequence (III) or (IV); (I) a computer representation of the amino acid sequences (III) and (IV); (J) a method comprising storing a computer representation of the amino acid sequences (III) and (IV).

USE/ADVANTAGE - The methods allow the rapid humanisation of antibodies and require less DNA synthesis than previous methods. The humanised antibodies have desired binding and other characteristics and activities but are less antigenic in humans than non-human antibodies.

Dwg.3/6

FS CPI EPI
FA Ab; GI
MC CPI: B04-B04C5; B04-B04C6; B04-C01G; D05-C11; D05-H11
EPI: T01-J06A

1123 ANSWER 65 CF '93 WPIX COPYRIGHT 2001 FERMENT INFORMATION LTD
AN 1992-3-5005 [47] WPIX
DNN N1992-3936A1 DMC C1992-11121
TI Correction of algorithm for identification of biological sample - involves comparing parameters with standard values, correcting membership characteristic **function** and identifying type of organism.

DC B04 D16 J04
FA (OMRON) OMRON CORP

CYC 1

PI JP 04283876 A 19921008 (199247)* 5P G06F015-62 <--
ADT JP 04283876 A JP 1991-48118 19910313

PPAI JP 1991-48118 19910313
IC ICM G06F015-62
ICS G01N033-48; G01N033-49

ICA G06F003-44

AB JP 04283876 A UPAB: 19931116
Biological samples including several organisms of two or more types are dried and stained after application, the images of the organisms are input as image signals, characteristic parameters for each organism are computed from the image signals, and the types of the organisms are identified using the characteristic parameters. The characteristic parameters obtnd. from the biological samples are compared with the characteristic parameters (standard values) used for the determin. of identification algorithm. According to the variation, the membership characteristic **function** of the identification algorithm is corrected, and the types of the organisms are identified using the corrected algorithm.

USE/ADVANTAGE - Used to correct the identification algorithm in classification of the types of organisms existing in a biological sample using the characteristic parameters based upon the morphological/ density (colour) characteristics of organisms (e.g., cell sorting, leukocyte sorting, etc.). The membership characteristic **function** of identification algorithm is corrected according to the variation between the characteristic parameters obtnd. from biological sample and those used for determining identification algorithm, so, even if the characteristic parameters vary, classification can be made with comparatively good accuracy.

3/5

Dwg.2/5

ES CPI
 FA AB; GI
 MC **EP1: B04-B04A; B04-B04D1; B11-B17B1; B12-K04;**
D05-H09; J04-B01

L123 ANSWER 66 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 1992-366389 [44] WPIX

CR 1994-130048 [17]; 1994-309404 [42]; 1991-169985 [22]; 1992-398889 [48];
 1994-109139 [17]; 1994-271749 [44]; 1995-105497 [14]; 1995-119549 [18];
 1995-115113 [30]; 1995-145111 [30]; 1995-169531 [22]; 1996-371955 [36];
 1996-433027 [43]; 1996-446406 [44]; 1997-051642 [45]; 1997-039915 [09];
 1997-53351 [44]; 1997-97141 [11]; 1999-010106 [40]; 2000-112971 [13]

DNN N1992-273206 1997-0102-160745

TI Identification of analyte compsn. of sample - comprising assessing sample w.r.t. survey parameter and classifying using SC **profiles**, for recognising herbicides, therapeutic use etc..

DC P 4 '07 DIC SOC

IN AMELER, S M; KAUVAR, L M

FA (TEFF-EU) TERRAFIN TECHNOLOGIES INC; (TELI-N) TELIK INC

CYC 21

PI WO 9217784 A1 19921015 (1993144)* EN 33p G01N033-563
 FW: AT BE CH DE DK ES FR GB GR IT LU MC NL SE
 W: AU CA HK JP KR RS
 AU 9219878 A 1994-1192 (199401); G01N033-563
 EP 581881 A1 19940109 (199401); EP G01N033-563
 E: AT BE CH DE DK ES FR GB GR IT LI LU MC NL SE
 US 5336659 A 19940316 (199432); 10p G12B20/1-00
 JP 06509865 W 19941107 (199401); 10p G01N033-563
 EP 581881 A4 19940706 (199432); G01N033-563
 AU 5336659 P 19940317 (199401); G01N033-563
 EP 581881 P1 19940306 (199401); EP G01N033-563
 E: AT BE CH DE DK ES FR GB GR IT LI LU MC NL SE
 DE 64226539 E 19920402 (199202) G01N033-563
 JP 3101323 B2 20001023 (200006) 12p G01N033-563

ADT WO 9217784 A1 WO 1992-032716 19920402; AU 9219878 A AU 1992-19878
 19920402, WO 1992-032716 19920402; EP 581881 A1 EP 1992-912207 19920402,
 WO 1992-032716 19920402; US 5336659 A US 1991-678349 19910401; JP 06509865
 W JP 1992-511681 19920402, WO 1992-032716 19920402; EP 581881 A4 EP
 1992-912207 ; AU 662001 B AU 1992-19878 19920402; EP 581881 B1 EP
 1992-912207 19920402, WO 1992-032716 19920402; DE 69218539 E DE
 1992-628539 19920402, EP 1992-912207 19920402, WO 1992-032716 19920402; JP
 3101323 B2 JP 1992-511681 19920402, WO 1992-032716 19920402

FET AU 9219878 A Based on WO 9217784; EP 581881 A1 Based on WO 9217784; JP
 06509865 W Based on WO 9217784; AU 662001 B Previous Publ. AU 9219878,
 Based on WO 9217784; EP 581881 B1 Based on WO 9217784; DE 69218539 E Based
 on EP 581881, Based on WO 9217784; JP 3101323 B2 Previous Publ. JP
 06509865, Based on WO 9217784

PFAI US 1991-678349 19910402

REP 1.Jnl.Ref; 2.Jnl.Ref; WO 8600991

IC ICM C12Q001-00; G01N033-53; G01N033-54; G01N033-56
 ICS G01N033-557; G01N033-566; G01N033-564; G01N033-54; G06F015-40
 ; G06F017-30

AB WO 9217784 A WFA: 20001106
 Determin. of the analyte compsn. of a sample contg. at least one member analyte of a cross-reactive set of analytes comprising: (a) assessing the sample w.r.t. each of n information channels in a panel contg. n information channels for characteristic parameters of the members of the set to obtain a survey of characteristics (E.g: **profile**; and (b) comparing the SC **profile** with a reference set of **profiles** obtnd. from known compsn. of the members of the cross-reactive set.

USE/ADVANTAGE - Analysis of samples for recognising classes of analytes, e.g., herbicides, in clinical analysis of biological samples (e.g., steroids, **protein** hormones or their metabolites), artificial materials designed for therapeutic use (e.g., beta-blockers, chemotherapeutic agents, prostaglandin, inhibitors, etc.). Illicit drugs

may be detected and distinguished for forensic purposes.

Dwg.1A'3

FC CPI EPI
 FA AB; GI; DCN
 MC CPI: B04-B02D; C04-B02D; B04-B02E; C04-B02E; B04-B04C5; C04-B04C5;
 B07-D013; C07-D13; B12-G01; C12-G01; B12-K04A; C12-K04A; B12-P05;
 C12-P05; D05-H04
 EPI: S03-E14H4

ABEQ US 5133854 A UFAB: 1994094/8

Identification of analyte compsns. is enhanced by applying pattern recognition techniques. Samples are reacted with binding reagents to produce a datatank of known concn. of each analyte, which is stored in a computer. Unknown samples are then matched against the patterns.

ADVANTAGE - Identification is precise.

Dwg.1a 3

1023 ANSWER 67 OF 73 WPIX © COPYRIGHT 1991 DERWENT INFORMATION LTD
 AN 1992-349151 [4]; WPIX
 CR 1992-415765 [50]; 1992-248451 [20];
 DUN N1992-266381 DNC C1992-155004
 TI New lactam derivs. cross-reactive with antikey CDR - useful for producing prophylactic or therapeutic immune response.
 DC P03 B04 S07
 IN GREENE, M I; KAHN, M; SARASOVI, H P
 FA UTPB-H; UNIV PENNSYLVANIA
 CYC 19
 PI WO 9216548 A1 19921021 (1992); EP 45p C07F005-02 <--
 FW: AT BE CH DE DK ES FR GB GR IT LU MC NL SE
 W: AU CA FR KR
 AU 9217731 A 19921021 (1992) C07F005-01 <--
 JP 06510741 W 19921021 (1992) 11p C07F005-01
 ALT WO 9216548 A1 WO 1992-082383 19920924; AU 9217731 A AU 1992-17731
 19920324, WO 1992-082383 19920924; JP 06510741 W JP 1992-509950 19920324,
 WO 1992-082383 19920324
 FDT AU 9217731 A Based on WO 9216548; JP 06510741 W Based on WO 9216548
 PPAI US 1991-674866 19910325
 REP S.Jnl.Ref; US 4939666; US 4946778
 IC ICM C07K005-02
 ICS A61K031-395; A61K039-395; C07P255-02; C07K005-12;
 C07K015-28; G06F003-14

AB WO 9216548 A UFAB: 19940921
 (I) is immunologically cross reactive with at least one complementarity determining region (CDR) of at least one antibody (Ab), the CDR having a plurality of hydroxyl gps. positioned in three-dimensional space, where (I) comprises the same number of hydroxyl gps. as the CDR and there exists at least one conformation of the synthetic cpd. where the 3-D positioning of its hydroxyl gps. is identical to ten 3-D positioning of the hydroxyl gps. of the CDR. The synthetic cpd. (I) is of formula: where R1, R2 and R3 = chemical **functional** gps. each comprising at least one hydroxyl gp.

USE/ADVANTAGE - (I') may be included in compsns. for admin. to mammals to elicit a prophylactic or therapeutic immune response. Methods are provided for the prepn. of synthetic cpds. which are immunologically cross reactive with individual Ab CDRs. These cpds. may be used in studying the role of a CDR chemical **functionalitY** to Ab or CDR activity e.g. in studying the role of specific aminoacid residues in the interaction of the monoclonal Ab 87.92.6 light chain variable region with the revirus type I receptor. Compsns. comprising relativley small synthetic cpds. may be developed to mediate the biological effects of Abs or other ligands. These cpds. should possess beneficial properties such as increased half-life, the ability to cross the blood-brain barrier, and lack of immunogenicity and thus shcould be useful for the development of new pharmaceutical, therapeutic, diagnostic and receptor binding agents.

5/5

Dwg.5/5

Dwg.5/5

FS CPI EPI
 FA AB; GI; DCN
 MC CPI: B02-V02; B07-D15; B07-D13; B11-C08; **B12-K04**
 EPI: SG3-E14H4

L123 ANSWER 69 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 1990-138364 [16] WPIX
 DNN C1990-060730
 TI Apps. to analyse and process **genome** - comprises a **genome** character sequence analyser with window **function** unit allowing character sequence extn.
 DC B04 D16
 PA (FUIT) FUJITSU LTD
 CYC 1
 PI JP 04075582 A 19900310 (199214)* 30
 JP 04075582 B2 19901118 (199351) 1sp G06F017-30 <--
 AUT JP 04075582 A JP 1990-188364 19900717; JP 0325948 B2 JP 1990-188364
 19900717
 FDT JP 04075582 B2 Previous Publ. JP 04075582
 PFAI JP 1990-188364 19900717
 IC C12M001-00
 ICM G06F017-30
 ICS C12M001-00
 AP JP 04075582 A UPAB: 19901006
 Apps. comprises a **genome** character sequence analyser with a window **function** unit to scan a character sequence. This specifies the **genome** to thereby extract the character sequence.
 USE - For use in biological laboratories. (0/0)
 0/0

FS CPI
 FA AE
 MC CPI: **B04-B04A1; B11-C08; B12-K04A; D05-H09**

L123 ANSWER 69 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 1992-009001 [02] WPIX
 DNN N1992-007078 DNN C1992-003939
 TI Obtaining signals representing different species in analysis - using ratios of signals obtd. in different manner and reacting differently to species, used to obtain DNA sequence information.
 DC B04 D16 J04 S03
 IN DAM, R J; DAVIS, J E; HOCHBERG, A M; PROBER, J M
 FA (DUPO) DU PONT DE NEMOURS & CO E I
 CYC 10
 PI EP 463591 A 19900102 (199202)* 24p
 F: DE FR GB IT NL SE
 CA 2045720 A 19911230 (199213)
 US 5119316 A 19920602 (199215) 1sp G01N030-00
 JP 05068598 A 19930303 (199316) 1sp C12Q001-68
 EP 463591 B1 19970924 (199743) EN 24p G01N027-447
 R: DE DK FR GB IT NL SE
 DE 69127721 E 19971030 (199749) G01N027-447
 JP 2993141 B2 19901220 (200005) 1sp C12Q001-68
 AUT EP 463591 A EP 1991-110342 19910627; US 5119316 A US 1990-545746 19900629;
 JP 05068598 A JP 1991-184041 19910628; EP 463591 B1 EP 1991-110342
 19910622; DE 69127721 E DE 1991-697721 19910622, EP 1991-110342 19910622;
 JP 2993141 E JP 1991-184041 19910626
 FDT DE 69127721 E Based on EP 463591; JP 2993141 B2 Previous Publ. JP 05068598
 PFAI US 1990-545746 19900629
 REP EP 157260; EP 198403; EP 242629; EP 294524; EP 294996
 IC ICM C12Q001-68; G01N027-447; G01N030-00
 ICS G01N021-64; G01N027-44; G06F015-353; G06F017-17
 ICA C12N015-00
 AB EP 463591 A UPAB: 19901006
 Two signals which react differently to different species, following species separation in time or space, are generated, each signal varying in amplitude in respective senses to form peaks as a **function** of

the nature of the species, each peak being representable by the sum of three or more peak templates. A template is derived for fitting to each of the signals to obtain the peak amplitude of the signals and a third signal is derived, from the ratio of a **function** of the peaks of the first and second signals, the amplitude of the third signal being indicative of the identity of each of the species.

USE/ADVANTAGE - Particular in obtaining DNA sequence information. Allows the underlying peak shape to be periodically redetermined and fitted to the data in a way that yields accurate ratios, even when the precise location of the peak is not known. The method contains self consistency checks so that potentially erroneous ratios, and hence base determinations may be recognised.

4A/4

FS WPI EPI
 FA AE; GI
 MC WPI: B04-B04A1; B11-C03; B12-K04; D05-H09;
 L05-H12; J04-E01
 EPI: S03-E03E

ABEQ US 5119316 A UPAE: 19-31066
 Determn. of DNA sequences comprises generating signal by a sequencer that varies in amplitude to form peaks as a **function** of the nature of the species and generating second signal varying in amplitude in a second sense different from the first. A fraction of the first signal is subtracted from the second signal, from the peak centre of the first peak to the peak centre of the second signal peak in each peak sequence, so amplitude of peaks in each difference is zero, giving a prototypical template for the peak. The template is fitted to the first and second signals to give a third signal indicative of the identity of each of the species.

ADVANTAGE - Improved accuracy of sequence information.

ABEQ EP 463831 B UPAE: 19-71031
 A method of identifying the sequence of bases in a particular DNA fragment by means of signals characteristic of reporter molecules associated with a particular base the method including (a) electrophoretically separating fragments in time and/or space, (b) for the particular fragment, producing an emission spectra which is passed through separate filters having different characteristics to generate a first and second signal, each signal varying in amplitude as a **function** of the reporters associated with the bases in the fragment, to form peaks which correspond to a particular reporter associated with a particular base, any peak of first and second signal being approximately representable by the sum of three or more peak templates, and (c) obtaining a third signal corresponding to the ratio of the first and second signals, the amplitude of the third signal being indicative of the identity of each of the bases, characterised in that the third signal is obtained by the steps: finding a sequence of peaks XYX in each of the first and second signals where X represents one base and Y represents another base, subtracting from the peak centre of the first X to the peak centre of the second X in each signal such that the amplitude of the X peaks in the difference is substantially zero, thereby obtaining a prototypical template for the peak of each signal, and fitting the respective prototypical template to the respective first and second signal spectrums to obtain the peak amplitude of the first and second signals for obtaining the third signal from the ratio of such peak amplitudes.

Dwg.0/4

L123 ANSWER 70 OF 73 WPIX COPYRIGHT 2001 DEWEY INFORMATION LTD
 AN 1991-193323 [26] WPIX
 DNN N1991-147485 SNC C1991-083716
 TI High resolution optical scanner for viewing object - uses flat support and linear optical sensor with drive mechanism to successively seal object.
 DC B04 D16 S03 S05
 IN EMERY, R J; HEMMINGES, F W; EMERGY, R J
 FA (USBI-N) US BIOCHEMICAL CORP; (USBI-N) US PICCHEM CORP
 CYC 26
 PI WO 9108465 A 19910613 (199126)*

PW: AT BE CH DE DK ES FR GB GR IT LU NL SE
 W: AU BG BE CA FI HU NL KP KR NO RO SU
 AU 9171671 A 19910626 (199111)
 US 5297188 A 19940312 (199411) 60p G06F009-44 <--
 ADT US 5297188 A Div ex US 1988-142553 19891128, Cont of US 1991-640471
 19910119, US 1991-640471 19910119
 PRA1 US 1991-64055; 19910119
 REP 1.Jnl.R-f; US 4717217; US 4961920
 IC G01N011-71
 ICM G06F009-44
 ICS G01N011-75; G06F013-10; G06F015-42;
 G06F015-70

AB WO 9118145 A (PAB: 19910119)
 The apparatus for optically scanning an object comprises a flat support for holding the object. A linear optical sensor is held in a fixed position on one side of support. A light source is held in a fixed position on the other side of the support for projecting light through the object toward the optical sensor. A drive mechanism moves the object along one direction to successive scanning positions. The user may create custom application computer programs for performing desired image analysis on an object.

The system stores software tools capable of performing an image scanning & analysis **function**, stores interpreter program corresponding to the natural language identifier for the **functions** and provides an interactive computer environment for user.

USE - Analysing autoradiographic image of products of deoxyribonucleic acid sequencing reaction.

17/17

FS C01 E01

FA AB

MC C01: B04-B04A1; B11-C01B1; B12-K04A; D05-H09
 ; D05-H11

E01: S03-E04; S03-E14H; S04-C09

ABEQ US 5297188 A (PAB: 19940318)
 A data processing system comprises a computer, an image scanner and a natural language interface responsive to user input. The system creates and executes a custom application programme with a sequence of user-specified software tools for scanning an image with four vertical sections, each with no or some horizontal bands with non-uniform perimeters.

Tools are stored for correcting scanning and analysis and custom programme is created by storing a user-defined sequence of tools using the natural language identifiers, and the programme is executed to scan the image and store the data, using the data to locate the horizontal bands and to correlate each of these to a predetermined shape.

USE/ADVANTAGE - Partic. for analysing DNA autoradiogram, for general imaging, populations screening and diagnosis including AIDS testing, and micro-titre plate reading. Provides high resolution analysis of a wide range of formats in an interactive computer environment.

Dwg.17A/17

L123 ANSWER 71 OF 75 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 1990-254121 [75] WPIX
 CR 1991-336074 [42]; 1994-0655-3 [06]; 1994-176282 [21]; 1999-454069 [33]
 DNN N1990-196001 LNC C1990-110079
 TI Screening for foetus with down syndrome - by measuring pregnant woman's blood levels of free beta sub-unit of human chorionic gonadotropin.
 DC B04 B16 C05 S05
 IN MACRI, J N
 PA (MACRI-I) MACRI J N
 CYC B3
 PI WO 9009305 A 19900706 (199003)* 50p
 PW: AT BE CH DE DK ES FR GB IT LU NL CA SE
 W: AU BB BR CA DK FI HU IP KP KR LK MD MG MW NO RO SD SU
 AU 4050336 A 19900813 (199044)
 EP 400056 A 19910130 (199105)

ABEQ US 5253917 A UPAB: 19931120

Appts. for determining if a pregnant woman is at significant risk of carrying a foetus with Down's syndrome measures the maternal serum level of free beta human chorionic gonadotrophin and has a computer to calculate a set of normative data and incorporate the measurement and the gestational age into a probability density **function**, thus comparing the actual measurement's with the normative data set.

The computer program generates the **function** by a linear discriminant analysis procedure. Measurements of intact human chorionic gonadotrophin, alpha-fetal **protein** and unconjugated oestriol may also be input into the computer for comparison with normative data.

ADVANTAGE - Correctly predicts a higher percentage of foetal Down's syndrome cases with a lower false positive rate than conventional methods.

Lwg.0/14

ABEQ US 5254618 A UPAB: 19940110

Screening pregnant women for a Down's syndrome foetus comprises measuring free beta-HCG and comparing this and the woman's gestational age to reference values in pregnant women with normal/Down's syndrome foetuses. Higher level of beta-HCG indicates a higher probability of Down's syndrome.

ADVANTAGE - Greater percentage of foetal Down's syndrome with a lesser false positive rate.

Lwg.1/12

ABEQ EP 400956 B UPAB: 19960418

An in vitro screening method for determining if a pregnant woman is carrying a fetus with Down syndrome comprising: assaying a pregnant woman's blood for free beta human chorionic gonadotropin (hCG), the results of the assay being indicative of increased risk of fetal Down syndrome.

Lwg.1/14

L123 ANSWER 72 OF 73 WPIX COPYRIGHT 2001 DEFVENT INFORMATION LTD

AN 1990-100003 [14] WPIX

DNN N1990-077625 DMC C1990-044237

TI Measurement of lipoprotein constituents of blood - by building up reference NMR spectra of individual constituents to match measured NMR spectrum of blood.

DC B04 S03 S05

IN OTVOS, J E

PA (OTOB-I) OTOBOS J A; (OTVO-I) OTVOS J D

CYC 13

PI EP 361214 A 19900404 (199014)* EN 45p

F: BE CH DE ES FR GB IT LI NL

JP 02116743 A 19900501 (199023)

US 4933844 A 19900612 (199031)

CA 1327993 C 19940322 (199417)

EP 361214 B1 19940713 (199427) EN 43p G01N024-08

F: BE CH DE ES FR GB IT LI NL

DE 68916729 E 19940818 (199431) G01F013-46

ES 2056173 T3 19941001 (199440) G01F013-46

JP 3059181 B2 20000704 (200016) 39p G01N014-00

ADT EP 361214 A EP 1989-116979 19890913; JP 02116743 A JP 1984-246216

19890920; US 4933844 A US 1986-148750 19880920; CA 1327993 C CA

1989-612978 19840912; EP 361214 B1 EP 1989-116979 19840912; DE 68916729 E

DE 1984-616729 19890313; EP 1984-116979 19890913; ES 2056173 T3 EP

1984-116979 19890913; JP 3059181 B2 JP 1989-246216 19840920

F1T DE 68916729 E Based on EP 361214; ES 2056173 T3 Based on EP 361214; JP 3059181 B2 Previous Publ. JP 02116743

PRA1 US 1988-243750 19890426

REP 5.Jnl.Ref; EP 234524; 4.Jnl.Ref

IC G01N024-08; G01N035-02; G01F033-46; G06F015-42

ICM G01N024-00; G01N014-04; G01F033-46

ICG A61F001-055; G01N037-01; G06F015-42

AB EP 361214 A UPAB: 19920326

NMR spectrum of each lipoprotein constituent of blood is stored as a reference spectrum for that constituent and the measured NMR spectrum of a

plasma or serum sample of blood to be analysed is compared with a calculated lineshape produced by adding together the reference spectrum for each constituent, in amts determined by respective constituent coeffs. which are adjusted until the calculated lineshape matches the measured NMR spectrum of the sample. The concn. of at least one lipoprotein constituent of the blood sample can be calculated as a **function** of the constituent coeff. USE/ADVANTAGE - In the measurement of cholesterol in blood samples. Constituents of blood VLDL, LDL, HDL, and **protein** have NMR spectral properties invariant from person to person, so any differences in NMR spectrum of blood samples are due to the differences in levels of the constituents.

The process is simple and reliable and provides accurate measurements of lipoprotein content of blood.

4/4

ES CPI 5H1
 FA AB; GI; DGN
 MC CPI: E01-C02; B04-B04D1; B04-B14L4; B11-C08A; **B12-K04A**
 EPI: S03-E07; S03-E14H1; S05-C01; S05-D02K
 ABEQ US 4933844 A UPAB: 1930928

The lipoprotein constituents of blood are measured by (i) storing the NMR spectrum of each lipoprotein constituent as a reference spectrum for that constituent; (ii) acquiring an NMR signal produced by plasma or serum sample of the blood to be analysed in an NMR spectrometer; (iii) producing an NMR spectrum of the sample by transforming the signal; (iv) producing a calculated lineshape by adding together the reference spectrum for each constituent in amts. determined by respective constituent coeffs.; (v) adjusting the coeffs. to fit the calculated lineshape to the NMR spectrum of the sample; and (vi) calculating the concn. of at least one lipoprotein constituent as a **function** of the value of its constituent coeff.

ADVANTAGE - A method for the accurate and reliable measurement of the lipoprotein constituents of blood is provided.

ABEQ EP 561114 B UPAB: 19340414

A method for measuring the lipoprotein constituents of blood, the steps comprising: storing the NMR spectrum of each lipoprotein constituent as a reference spectrum for that constituent; acquiring the NMR spectrum of a plasma or serum sample of blood to be analysed; producing a calculated line shape by adding together the reference spectrum for each constituent in amounts determined by respective constituent coefficients; adjusting the constituent coefficients to fit the calculated line shape to the NMR spectrum of the sample; and calculating the concentration of at least one lipoprotein constituent as a **function** of the value of its constituent coefficient.

Dwg.0/4

L123 ANSWER 73 OF 73 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD

AN 1986-286053 [44] WPIX

DNN N1986-213729 DMC C1986-123760

TI Determining base sequence of nucleic acid - by subjecting digital signals from auto-radiograph to signal processing.

DC B04 D16 K08 S03

IN HARA, M

PA (FUJIF) FUJI PHOTO FILM CO LTD

CYC 5

PI EP 199327 A 19861029 (198644)* EN 36p

E: DE FR NL

JF 61243361 A 19861029 (198650)

JF 61243361 A 19861029 (198650)

US 4720786 A 19860113 (198605)

EP 199327 B 19920904 (199210)

E: DE FR NL

DE 3684030 G 19920409 (199216)

JF 94061341 B 19921006 (199214) 10p G01N33-60

JF 94061342 B 19921006 (199214) 10p G01N33-60

ADT EP 199327 A EP 1986-105493 19860421; JP 61243361 A JP 1986-65275 19850419; JP 61243361 A JP 1985-65276 19850419; US 4720786 A US 1986-654381

19860421; JP 04062341 B JP 1985-85275 19850419; JP 04062342 B JP 1985-85276 19850419

FDT JP 04062341 P Based on JP 61243361; JP 04062342 B Based on JP 61243362

PRAI JP 1985-85275 19850419; JP 1985-85276 19850419

REP 1.Jnl.Prf; A2...8939; EP 132621; CP 59776754; No-SP.Pub

IC ICM G01N33/60

ICS Cl.0901-6; G01N27-447; G01N33-60; G01T91-20; G06F015-20

AB EP 199327 A UPAB: 19930921

A signal processing method for determining base sequence of nucleic acids comprises subjecting digital signals to signal processing. The digital signals corresp. to an autoradiograph of resolved rows which are formed by resolving a mixt. of base-specific DNA fragments or base-specific RNA fragments labelled with a radioactive element in one-dimensional direction on a support medium.

Method comprises (a) detecting at least 2 bands in the lower part of each resolved row and numbering the bands consecutively from the lower end, (b) obtaining correlation between the band's number and a resolved distance for each resolved row and (c) determining difference in the resolved distance between the resolved rows from the resulting correlation and making correction for resolved position on each row by taking the difference as locational deviation of the rows from each other.

USE/ADVANTAGE - The base sequence of the nucleic acids can be simply determined with high accuracy by processing the digital signals through a signal processing circuit having a **function** capable of making the correction for offset distortion even when the resolved pattern causes the offset distortion.

0/6

FS CPI EPI

FA AB

MC CPI: B04-B04A1; B11-C07B5; B12-K04E; E05-H12; F03-B;
KG9-E

EPI: SG3-E14H

ABEQ DE 3664030 G UPAB: 19930921

A signal processing method for determining base sequence of nucleic acids comprises subjecting digital signals to signal processing. The digital signals corresp. to an autoradiograph of resolved rows which are formed by resolving a mixt. of base-specific DNA fragments or base-specific RNA fragments labelled with a radioactive element in one-dimensional direction on a support medium.

Method comprises (a) detecting at least 2 bands in the lower part of each resolved row and numbering the bands consecutively from the lower end, (b) obtaining correlation between the band's number and a resolved distance for each resolved row and (c) determining difference in the resolved distance between the resolved rows from the resulting correlation and making correction for resolved position on each row by taking the difference as locational deviation of the rows from each other.

USE/ADVANTAGE - The base sequence of the nucleic acids can be simply determined with high accuracy by processing the digital signals through a signal processing circuit having a **function** capable of making the correction for offset distortion even when the resolved pattern causes the offset distortion. ()

ABEQ EP 199327 P UPAB: 19930922

A signal processing method for determining the base sequence of nucleic acids by subjecting digital signals (DxBy) to signal processing, said digital signals (DxBy) corresponding to an autoradiograph of plural resolved rows which are formed by resolving a mixture of base-specific DNA fragments or base-specific RNA fragments in one-dimensional direction on a support medium, to form bands (n), said fragments being labelled with a radioactive element, said method comprising the steps of: (1) detecting at least two bands (n) in the lower part of each resolved row, and numbering all the bands (n) consecutively from the lower end, (2) obtaining a correlation between the number of which band (n) and the corresponding migration distance (y') for each resolved row, (3) determining the difference in the migration distance (y') between the resolved rows from the resulting correlations, and making corrections for the migration positions on each row by taking said differences as the locational

deviation of the rows from each other.

ABEQ US 4720746 A UPAB: 19950922

Signal processing method for nucleic acid base sequence determin. is applied to digital signals corresp. to an autoradiograph of resolved rows formed by resolving a mix. of base specific DNA or RNA fragments labelled with radioactive element in one dimensional direction on a support medium.

At least two bands in the lower part of each resolved row are detected and numbered from the lower end. Correlation between band number and resolved distance is obtai. from each resolved row. Difference in resolved distance between the resolved rows from the resulting correlation is determined and correction is made for the resolved position of each row by taking the difference as location deviation of the rows from each other.

ADVANTAGE - Method is automatic and highly accurate.

→ d all abeq tech tot

L127 ANSWER 1 OF 2 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
 AN 1991-071531 (06) WPIX
 DNN HJ031-064116 PNC C0001-000116
 TI Methods, computer programs and databases for analyzing and make use of gene haplotype information.
 DC E04 D01 S03 T01
 IN DENTON, P E; JUJON, R S; FUANG, G; STEPHENS, J C; WINDEMUTH, A F; XU, C
 PA (GENA-A) RENAISSANCE PHARM INC
 CYC 94
 PI WO 2001001118 A2 20011104 200104 1 EN 277p G-6F 01-00 --
 FW: AT BE CH CY DE DK EA ES FI FF GB GH GM GA IE IT FR IS LU MC MW MZ
 CL DA PT JP SE SL SU TZ TR UW
 W: AE AG AL AM AT AU AV BA BB BG BS BY ES CA CH CN CP CU CZ DE DK DM
 DE EE ES FI GB GD GE GH GM HR HW ID IL IN IS JP KE FG KP FR KZ LC
 LK LR LS LT LU LV MA MD MG MK MN MW MX MC NO NZ PL PT RO RU SD SE
 SG SI SK SL TJ TM TR TT TC UA UG US US VI YU CA SW
 ADT WO 2001001118 A2 WO 2000-031754C 20000626
 PRAI US 1993-141521 19930625
 IC ICM G06F000-00
 AB WO 200101218 A2 UPAB: 200101-07
 NOVELTY - Methods, computer programs and databases for analyzing and make use of gene haplotype (HT) information, e.g. to determine the frequency of HTs in a population, to find correlations between HTs or genotype and a clinical outcome and to predict HTs from an genotype for a gene, are new.
 DETAILED DESCRIPTION - INDEPENDENT CLAIMS are included for the following:
 (i) a method (M) for generating a HT database for a population, comprising data elements representative of the HTs for at least 1 locus from the individuals (indivs) in the data base;
 (2) a (M) of predicting the presence of a HT pair in an indiv;
 (3) a (M) for identifying a correlation between a HT pair and a clinical response to a treatment, or other phenotype Pt.;
 (4) a (M) for identifying a correlation between a HT pair and a susceptibility to a condition or a disease of interest (OI), or other Pt. (OI);
 (5) a (M) of predicting an indivs response to a medical or pharmaceutical treatment;
 (6) a computer implemented (C-I) (M) for generating a gene structure screen for display on a display device (DD);
 (7) a C-I (M) for generating a HT pair frequency screen for display on a DD;
 (8) a C-I (M) for generating a linkage screen for display on a DD;
 (9) a C-I (M) for generating a phylogenetic tree screen for display on a DD;
 (10) a C-I (M) for generating a genotype (Gt.) analysis screen for display on a DD;
 (11) a (M) of displaying clinical response values of a subject

population as a function of HT pairs of the indivs in the population;

(12) a C-I (M) for carrying out a genetic algorithm for finding an optimal set of weights to fit a function of polymorphic site data to a clinical response measurement;

(13) a C-I (M) for displaying correlations between clinical outcome values for a selected population;

(14) a (M) for conducting a clinical trial of a treatment protocol for a medical condition (OI);

(15) a (M) of inferring Gts. of indiv subjects for a selected gene having polymorphic sites;

(16) a (M) of determining polymorphic sites or sub-HTs that correlate with a clinical response or outcome (OI);

(17) a (M) of determining polymorphic sites or sub-HTs that correlate with a clinical response or outcome (CI);

(18) a computer readable (C-U) medium (Md.) having computer readable (C-R) program code (PC) stored upon it, for causing a computer (Comp.) to adjust observed HT pair frequencies within a population group (the HT pair frequencies are stored in a C-R database of HT information for a gene or gene feature (OI));

(19) a C-U Md. having C-R PC stored upon it, for causing HT pair assignments to be made to an indiv member of a population whose Gt. information for a gene feature (OI) is stored in a C-R form;

(20) a C-U Md. having C-R PC stored upon it, for causing a Comp. to identify a correlation between a clinical response to a treatment or other Pt. and a HT or HT pair present at a candidate locus associated with the clinical response or other Pt.;

(21) a C-U Md. having C-R PC stored upon it, for causing a Comp. to identify a correlation between an indiv's susceptibility to a condition or disease (OI) or other Pt., and a HT or HT pair present at a candidate locus associated with the susceptibility to the condition or disease (OI) other Pt. (OI);

(22) a C-U Md. having C-R PC stored upon it, for causing a Comp. to predict an indivs response to a medical or pharmaceutical treatment based on one or more selected HTs or HT pairs of the indiv;

(23) a C-U Md. having C-R PC stored upon it, for causing a Comp. to display a gene's structure and gene features on a display device (D);

(24) a C-U Md. having C-R PC stored upon it, for causing a Comp. to display on a DD, HT frequency data within a population of indivs, for a selected gene or gene feature;

(25) a C-U Md. having C-R PC stored upon it, for causing a Comp. to display on a DD, polymorphic site linkage data for a gene or gene (OI);

(26) a C-U Md. having C-R PC stored upon it, for causing a Comp. to display on a DD a **phylogenetic tree**;

(27) a C-U Md. having C-R PC stored upon it, for causing a Comp. to display a Gt. analysis screen on a DD;

(28) a C-U Md. having C-R PC stored upon it, for causing a Comp. to display clinical response values, or other Pt. data, of a subject population as a function of HT pairs of the indivs in the population;

(29) a C-U Md. having C-R PC stored upon it, for causing a Comp. to display on a DD, clinical response values, or other Pt. data, of a subject population as a function of HT pairs of the indivs in the population, for a gene or gene feature (OI);

(30) a C-U Md. having C-R PC stored upon it, for causing a Comp. to carry out a genetic algorithm for finding an optimal set of weights to fit a function of polymorphic site data for a gene or gene feature (OI) to a clinical response measurement;

(31) a C-U Md. having C-R PC stored upon it, for causing a Comp. to display on a DD, correlation between clinical outcome values obtained from selected clinical outcome measures for a selected population;

(32) a C-U Md. having C-R PC stored upon it, for causing a Comp. to provide information of use in conducting clinical trials of a treatment protocol for a medical condition (OI);

(33) a C-U Md. having C-R PC stored upon it, for causing a Comp. to infer Gts. of indiv subjects for a selected gene having polymorphic sites;

(34) C-U media having C-R PC stored upon it, for causing a Comp. to determine polymorphic sites or sub-HTs that correlate with a clinical

response or outcome (O1), or other Pt. (O1);

(35) Comps. programmed to carry out the above (Ms) or comprising the above Comp.-useable or -readable media, comprising a memory with at least 1 region for storing Comp. executable PCs and a processor for executing the PC stored in the memory;

(36) a data structure for storing an organizing biological information, stored on a C-R Md. and accessible by a processor, which comprises a single parent table which is adapted for storing, organizing and retrieving a number of genetic features by the relative positional relationships between the genetic features;

(37) a (M) for storing and organizing biological information; and

(38) a data structure for storing an organizing biological information, stored on a C-F Md. and accessible by a processor, which comprises a least 2 different fields, one of which includes a number of genetic features, and the other of which includes relative positional relationships between the genetic features.

Note: Further details of the above are given in the specification but had to be omitted from this abstract due to insufficient space.

USE - The methods, computer programs and databases for analyzing and make use of gene haplotype HT information, e.g. to determine the frequency of HTs in a population, to find correlations between HTs or genotypes and a clinical outcome or the effects of a therapeutic intervention and/or to predict HTs from an individual's genotype for a gene.

Dwg.0/49

FS CPI EPI

FA AB

MC CPI: B04-E01; B11-C08E; B11-C09; B12-K04AB; D05-H01; D05-H03; D05-H12; D05-H18; D05-J
EPI: S03-E14H; T01-J06A1; T01-S03

L127 ANSWER 2 - F 0 WPIX COPYRIGHT 2000 DEPARTMENT INFORMATION LTD

AN 2000-026642 [03] WPIX

DNN H2000-019943 DNC C2000-006875

TI Transition/transversion ratio estimation method - for purine, pyridine in DNA involves creating a taxonomic tree from input base sequence data.

DC B04 D16 T01

PA (SEIBUTSU BUNSHI KOGAKU KENKYUSHO KK

CYC 1

PI JP 11259530 A 19990924 (200003)* JA 28p G06F017-30 <--

ADT JP 11259530 A JP 1998-307740 19980925

PEAI JP 1998-35313 19980112

IC ICM G06F017-30

ICS C12N015-00

AB JP 11259530 A UPAB: 20000112

NOVELTY - The base substitution number assessed from type of transition-transversion is to create a taxonomic tree. The branch lengths of taxonomic tree are supposed. The presumed type of transition/transversion ratio is applied and assessed value of transition/transversion ratio is output. The base substitution number is assessed using a predefined equation that relates to compared sequence number.

DETAILED DESCRIPTION - The base sequence data with more than three sequences are input to data processor.

An INDEPENDENT CLAIM is also included for creating a phylogenetic tree from the actual base sequence data. Repeated calculations are done using the maximum evolution distance and the results are converged by a specified method, to obtain a right tree form.

USE - For estimation of transition/transversion ratio of components such as purine, pyridine in deoxyribonucleic acid (DNA) using their base sequence data e.g. of mitochondria of human and chimpanzee (claimed).

ADVANTAGE - The transition/transversion ratio in controlled domains such as human of mitochondrial DNA is assessed correctly and the taxonomic tree is also built correctly.

Dwg.0/6

FS CPI EPI

FA AB

IC: CPI: B04-E01; B06-D09; B07-D12; B11-C08E; B12-F04E; D05-H09
 EPI: T01-J05B

=> fil medline

FILE 'MEDLINE' ENTERED AT 14:57:21 ON 12 FEB 2001

FILE LAST UPDATED: 27 OCT 2000 (20001027/UP). FILE COVERS 1960 TO DATE.

MEDLINE thesauri in the /CN, /CT, and /MN fields incorporate the MeSH 2000 vocabulary. Enter HELP THESAURUS for details.

The OLDMEDLINE file segment now contains data from 1918 through 1965. Enter HELP CONTENT for details.

Left, right, and simultaneous left and right truncation are available in the Basic Index. See HELP SFIELDS for details.

THIS FILE CONTAINS CAS REGISTRY NUMBERS FOR EASY AND ACCURATE SUBSTANCE IDENTIFICATION.

MEDLINE UPDATES ARE ON HOLD UNTIL AFTER THE ANNUAL RELOAD HAS BEEN COMPLETED. NOTICE WILL BE GIVEN ONCE THE RELOAD IS COMPLETED AND RELOAD DETAILS WILL BE FOUND IN HELP RELOAD.

=> d all tot

L157 ANSWER 1 OF 23 MEDLINE
 AN 199903436 MEDLINE
 DN 9-083436
 TI B30.2-like domain proteins: update and new insights into a rapidly expanding family of proteins.
 AU Henry C; Mather I H; McDermott M F; Pontarotti P
 CS Centre National de la Recherche Scientifique URA 1485, Faculte de Medicine, Limoges, France.
 SO MOLECULAR BIOLOGY AND EVOLUTION, (1998 Dec) 15 (12) 1696-705.
 Journal code: MOB. ISSN: 0737-4038.
 CY United States
 DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 FS Priority Journals
 EM 199903
 EW 19990304
 AB The B30.2 domain is a conserved region of around 170 amino acids associated with several different protein domains, including the immunoglobulin folds of butyrophilin and the RING finger domain of ret finger protein. We recently reported several novel members of this family as well as previously undescribed protein families possessing the B30.2 domain. Many proteins have subsequently been found to possess this domain, including pyrin/marenostrin and the midline 1 (MID1) protein. Mutations in the B30.2 domain of pyrin/marenostrin are implicated in familial Mediterranean fever, and partial loss of the B30.2 domain of MID1 is responsible for Opitz G/BBB syndrome, characterized by developmental midline defects. In this study, we scrutinized the available sequence databases for the identification of novel B30.2 domain proteins using highly sensitive database-searching tools. In addition, we discuss the chromosomal localization of genes in the B30.2 family, since the encoded proteins are likely to be involved in other forms of periodic fever, autoimmune, and genetic diseases.
 CT Check Tags: Animal; Human; Support, Non-U.S. Gov't
 Amino Acid Sequence
 Autoimmune Diseases: GE, genetics
 Cattle
 Chromosome Mapping

Databases***Evolution, Molecular*****Hereditary Diseases: GE, genetics**
Molecular Sequence Data***Multigene Family*****Phylogeny*****Proteins: CH, chemistry*****Proteins: GE, genetics****Sequence Alignment****Sequence Homology, Amino Acid****Sinc Fingers**

CN 0 (Proteins)

L157 ANSWER 2 OF 23 MEDLINE

AN 109069409 MEDLINE

DN 99069409

TI Three novel proteins of the syntaxin/SNAP-25 family.

AU Steegmaier M; Yang B; Yoo J S; Huang B; Shen M; Yu S; Luo Y; Scheller R H
CS Department of Molecular and Cellular Physiology, Howard Hughes Medical
Institute, Stanford University School of Medicine, Stanford, California
94305-5345, USA.

JO JOURNAL OF BIOLOGICAL CHEMISTRY, (1998 Dec 18) 273 (51) 14171-9.

Journal code: JBI. ISSN: 0021-9258.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals; Cancer Journals

OS GENBANK-AF115323; GENBANK-AF115415; GENBANK-AF115436

EM 10906903

AB Intracellular membrane traffic is thought to be regulated in part by soluble N-ethylmaleimide-sensitive factor-attachment protein receptors (SNAREs) through the formation of complexes between these proteins present on vesicle and target membranes. All known SNARE-mediated fusion events involve members of the syntaxin and vesicle-associated membrane protein families. The diversity of mammalian membrane compartments predicts the existence of a large number of different syntaxin and vesicle-associated membrane protein genes. To further investigate the spectrum of SNAREs and their roles in membrane trafficking we characterized three novel members of the syntaxin and SNAP-25 (synaptosome-associated protein of 25 kDa) subfamilies. The proteins are broadly expressed, suggesting a general role in vesicle trafficking, and localize to distinct membrane compartments. Syntaxin 8 co-localizes with markers of the endoplasmic reticulum. Syntaxin 17, a divergent member of the syntaxin family, partially overlaps with endoplasmic reticulum markers, and SNAP-29 is broadly localized on multiple membranes. SNAP-29 does not contain a predicted membrane anchor characteristic of other SNAREs. In vitro studies established that SNAP-29 is capable of binding to a broad range of syntaxins.

CT Check Tags: Animal; Human; Support, Non-U.S. Gov't

Amino Acid Sequence

Blotting, Northern

Cell Line

COS Cells

Databases, Factual

Endoplasmic Reticulum: ME, metabolism

Ethylmaleimide: PD, pharmacology

Evolution, Molecular***Membrane Proteins: CH, chemistry*****Membrane Proteins: GE, genetics****Membrane Proteins: IP, isolation & purification****Molecular Sequence Data****Multigene Family*****Nerve Tissue Proteins: CH, chemistry*****Nerve Tissue Proteins: GE, genetics****Nerve Tissue Proteins: ME, metabolism****Phylogeny****Rats**

Recombinant Proteins: BI, biosynthesis
 Recombinant Proteins: CH, chemistry
 Sequence Alignment
 Sequence Homology, Amino Acid
 Transfection

RN 122-53-1 (Ethylmaleimide); 157546-56-0 (syntaxin)
 CN 0 (synaptosomal-associated protein 25); 0 (Membrane Proteins); 0 (Nerve Tissue Proteins); 0 (Recombinant Proteins)

L157 ANSWER 4 OF 23 MEDLINE
 AN 1949057494 MEDLINE
 DN 994907094

TI Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods.

AU Park J; Karplus K; Barrett C; Hughey R; Haussler D; Hubbard T; Chothia C
 CS MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK.
 SO JOURNAL OF MOLECULAR BIOLOGY, (1998 Dec 11) 284 (4) 1201-10.
 Journal code: JMB. ISSN: 0022-2836.

CY ENGLAND: United Kingdom

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals; Cancer Journals

EM 194903

EW 19490303

AB The sequences of related proteins can diverge beyond the point where their relationship can be recognised by pairwise sequence comparisons. In attempts to overcome this limitation, methods have been developed that use as a query, not a single sequence, but sets of related sequences or a representation of the characteristics shared by related sequences. Here we describe an assessment of three of these methods: the SAM-T98 implementation of a hidden Markov model procedure; PSI-BLAST; and the intermediate sequence search (ISS) procedure. We determined the extent to which these procedures can detect evolutionary relationships between the members of the sequence database PDBD40-J. This database, derived from the structural classification of proteins (SCOP), contains the sequences of proteins of known structure whose sequence identities with each other are 40% or less. The evolutionary relationships that exist between those that have low sequence identities were found by the examination of their structural details and, in many cases, their functional features. For nine false positive predictions out of a possible 432,680, i.e. at a false positive rate of about 1/50,000, SAM-T98 found 35% of the true homologous relationships in PDBD40-J, whilst PSI-BLAST found 30% and ISS found 25%. Overall, this is about twice the number of PDBD40-J relations that can be detected by the pairwise comparison procedures FASTA (17%) and GAP-BLAST (15%). For distantly related sequences in PDBD40-J, those pairs whose sequence identity is less than 30%, SAM-T98 and PSI-BLAST detect three times the number of relationships found by the pairwise methods. Copyright 1998 Academic Press.

CT Check Tags: Comparative Study; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.

Databases, Factual

Evaluation Studies

Evolution, Molecular

Markov Chains

Proteins: CH, chemistry

Proteins: GE, genetics

*Sequence Alignment: MT, methods

Sequence Alignment: SN, statistics & numerical data

Sequence Homology, Amino Acid

CN 0 (Proteins)

L157 ANSWER 4 OF 23 MEDLINE
 AN 1948417653 MEDLINE
 DN 98417653

TI Supersites within superfolds. Binding site similarity in the absence of homology.

AU Russell R B; Cacioli P D; Sternberg M J E
 CC Biomolecular Modelling Laboratory, Lincoln's Inn Fields, PO Box 123,
 London WC2A 3PX, UK.
 SO JOURNAL OF MOLECULAR BIOLOGY, (1998 Oct 2) 282 (4) 903-18.
 Journal code: JMB. ISSN: 0022-2836.
 CY ENGLAND: United Kingdom
 DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 FS Priority Journals; Cancer Journals
 EM 199809
 EW 19980903
 AB A method is presented to assess the significance of binding site
 similarities within superimposed protein three-dimensional (3D) structures
 and applied to all similar structures in the Protein Data Bank. For
 similarities between 3D structures lacking significant sequence
 similarity, the important distinction was made between remote homology (an
 ancient common ancestor) and analogy (likely convergence to a folding
 motif) according to the structural classification of proteins (SCOP)
 database. Supersites were defined as structural locations on groups of
 analogous proteins (i.e. superfolds) showing a statistically significant
 tendency to bind substrates despite little evidence of a common ancestor
 for the proteins considered. We identify three potentially new superfolds
 containing supersites: ferredoxin-like folds, four-helical bundles and
 double-stranded beta-barrels. In addition, the method quantifies binding
 site similarities within homologous proteins and previously identified
 supersites such as that found in the beta/alpha (TIM) barrels. For the
 nine superfolds, the accuracy of predictions of binding site locations is
 assessed. Implications for protein evolution, and the prediction of
 protein function either through fold recognition or tertiary structure
 comparison, are discussed. Copyright 1998 Academic Press.
 CT Check Tags: Animal; Human; Support, Non-U.S. Gov't
 *Binding Sites
 *Databases, Factual
 Evolution, Molecular
 Ferredoxins: ME, metabolism
 Ligands
 Models, Molecular
 Protein Conformation
 Protein Folding
 Protein Structure, Secondary
 *Proteins: CH, chemistry
 Proteins: ME, metabolism
 Sequence Alignment
 Sequence Homology, Amino Acid
 Statistical Distributions
 Structure-Activity Relationship
 CN 0 (Ferredoxins); 0 (Ligands); 0 (Proteins)

L157 ANSWER 5 OF 27 MEDLINE
 AN 1998391734 MEDLINE
 DN 99391734
 TI p50uASL, a 3'-5' oligoadenylylate synthetase like protein: a novel human
 gene related to the 3'-5' oligoadenylylate synthetase family.
 AU Hartmann E; Olsen H J; Widder S; Jorgensen R; Justesen J
 CC Department of Molecular and Structural Biology, University of Aarhus, C.
 F. Mollers alle Building 139, DK-4000 Aarhus C, Denmark and Human Genome
 Sciences Inc., 9410 Key West Avenue, Rockville, MD 20850, USA.
 SO NUCLEIC ACIDS RESEARCH, (1998 Sep 15) 26 (18) 4121-8.
 Journal code: NAR. ISSN: 0305-1048.
 CY ENGLAND: United Kingdom
 DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 FS Priority Journals; Cancer Journals
 GS GENBANK-AJ225089; GENBANK-AJ225390
 EM 199801
 EW 19990104

have evolved from the same RNA-binding proto-protein, which had a form of beta-loop-beta RNA binding motif.

CT Check Tags: Human; Support, U.S. Gov't, Non-P.H.S.

Amino Acid Sequence

Bacteria

Binding Sites

Computer Graphics

Computer Simulation

DNA-Binding Proteins: CH, chemistry

DNA-Binding Proteins: GE, genetics

DNA-Binding Proteins: ME, metabolism

*Evolution, Molecular

Models, Genetic

Models, Molecular

Molecular Sequence Data

Protein Folding

*Protein Structure, Secondary

*RNA-Binding Proteins: CH, chemistry

*RNA-Binding Proteins: GE, genetics

RNA-Binding Proteins: ME, metabolism

Sequence Alignment

Sequence Homology, Amino Acid

CN 0 (DNA-Binding Proteins); 0 (RNA-Binding Proteins)

LI57 ANSWER 7 OF 23 MEDLINE

AN 1998362531 MEDLINE

DN 98362531

TI Protein disorder and the evolution of molecular recognition: theory, predictions and observations.

AU Dunker A K; Garner E; Guillot S; Romero P; Albrecht K; Hart J; Obradovic Z; Kissinger C; Villafranca J E

CS Department of Biochemistry & Biophysics, Washington State University, Pullman 99164-4660, USA.. dunker@mail.wsu.edu

SO PACIFIC SYMPOSIUM ON BIOCOMPUTING, (1998) 473-84.
Journal code: CWQ.

CY Singapore

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

EM 199901

AB Observations going back more than 20 years show that regions in proteins with disordered backbones can play roles in their binding to other molecules; typically, the disordered regions become ordered upon complex formation. Thought-experiments with Schulz Diagrams, which are defined herein, suggest that disorder-to-order transitions are required for natural selection to operate separately on affinity and specificity. Separation of affinity and specificity may be essential for fine-tuning the molecular interaction networks that comprise the living state. For low affinity, high specificity interactions, our analysis suggests that natural selection would parse the amino acids conferring flexibility in the unbound state from those conferring specificity in the bound state. For high affinity, low specificity or for high affinity, multiple specificity interactions, our analysis suggests that the disorder-to-order transitions enable alternative packing interactions between side chains to accommodate the different binding targets. Disorder-to-order transitions upon binding also have significant kinetic implications as well, by having complex effects on both on- and off-rates. Current data are insufficient to decide on these proposals, but sequence and structure analysis on two examples support further investigations of the role of disorder-to-order transitions upon binding.

CT

Algorithms

Amino Acid Sequence

Binding Sites

Calmodulin: CH, chemistry

*Evolution, Molecular

Macromolecular Systems

*Protein Conformation
 Protein Structure, Secondary
***Proteins: CH, chemistry**
 Proteins: GE, genetics
 Selection (Genetics)
 Sequence Alignment
 Sequence Homology, Amino Acid
***Software**

CN 0 (Calmodulin); 0 (Macromolecular Systems); 0 (Proteins)

L157 ANSWER 3 OF 23 MEDLINE

AN 1998348042 MEDLINE

DN 98348042

TI Detecting distant relatives of mammalian LPS-binding and lipid transport proteins.

AU Beamer L J; Fischer D; Eisenberg D

CS Biochemistry Department, University of Missouri-Columbia, 65211, USA.

SO PROTEIN SCIENCE, (1998 Jul) 7 (7) 1643-6.

Journal code: BW. ISSN: 0961-8368.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

EM 199812

EW 19981_04

AB In mammals, a family of four lipid binding proteins has been previously defined that includes two lipopolysaccharide binding proteins and two lipid transfer proteins. The first member of this family to have its three-dimensional structure determined is bactericidal/permeability-increasing protein (BPI). Using both the sequence and structure of BPI, along with recently developed sequence-sequence and sequence-structure similarity search methods, we have identified 13 distant members of the family in a diverse set of eukaryotes, including rat, chicken, *Caenorhabditis elegans*, and *Biomphalaria galbrata*. Although the sequence similarity between these 13 new members and any of the 4 original members of the BPI family is well below the "twilight zone," their high sequence-structure compatibility with BPI indicates they are likely to share its fold. These findings broaden the BPI family to include a member found in retina and brain, and suggest that a primitive member may have contained only one of the two similar domains of BPI.

CT Check Tags: Animal; Human; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.

Biomphalaria

Blood Bactericidal Activity

***Blood Proteins: CH, chemistry**

***Blood Proteins: GE, genetics**

Blood Proteins: ME, metabolism

Caenorhabditis elegans

***Carrier Proteins: CH, chemistry**

Carrier Proteins: GE, genetics

Carrier Proteins: ME, metabolism

Chickens

Databases, Factual

Disulfides

Evolution, Molecular

Lipopolysaccharides: ME, metabolism

Molecular Sequence Data

Nerve Tissue Proteins: CH, chemistry

Nerve Tissue Proteins: GE, genetics

Protein Folding

Protein Structure, Tertiary

***Proteins: CH, chemistry**

Rats

Sequence Alignment

Sequence Homology, Amino Acid

Sequence Homology, Nucleic Acid

CN 0 (Bactericidal permeability increasing protein); 0 (cholesterol ester transfer proteins); 0 (lipopolysaccharide-binding protein); 0 (Blood Proteins); 0 (Carrier Proteins); 0 (Disulfides); 0 (Lipopopolysaccharides); 0 (Nerve Tissue Proteins); 0 (Proteins); 0 (TENP protein)

L157 ANSWER 9 OF 23 MEDLINE

AN 1996348036 MEDLINE

DN 96348036

TI The discoidin domain family revisited: new members from prokaryotes and a homology-based fold prediction.

AU Baumgärtner J; Hofmann K; Chiquet-Ehrismann R; Bucher P

CS Lund University, Department of Cell & Molecular Biology, Sweden... Stefan.Baumgärtner@medkem.lu.se

SO PROTEIN SCIENCE, (1998 Jul) 7 (7) 1626-31.

Journal code: BNW. ISSN: 0881-3204.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

EM 199812

EW 19981204

AB Members of the discoidin (DS) domain family, which includes the C1 and C2 repeats of blood coagulation factors V and VIII, occur in a great variety of eukaryotic proteins, most of which have been implicated in cell-adhesion or developmental processes. So far, no three-dimensional structure of a known example of this extracellular module has been determined, limiting the usefulness of identifying a new sequence as member of this family. Here, we present results of a recent search of the protein sequence database for new DS domains using generalized profiles, a sensitive multiple alignment-based search technique. Several previously unrecognized DS domains could be identified by this method, including the first examples from prokaryotic species. More importantly, we present statistical, structural, and functional evidence that the D1 domain of galactose oxidase whose three-dimensional structure has been determined at 1.7 Å resolution, is a distant member of this family. Taken together, these findings significantly expand the concept of the DS domain, by extending its taxonomic range and by implying a fold prediction for all its members. The proposed alignment with the galactose oxidase sequence makes it possible to construct homology-based three-dimensional models for the most interesting examples, as illustrated by an accompanying paper on the C1 and C2 domains of factor V.

CT Check Tags: Support, Non-U.S. Gov't
Amino Acid Sequence

*Bacterial Proteins: CH, chemistry

Bacterial Proteins: CL, classification

Databases, Factual

Evolution, Molecular

*Fungal Proteins: CH, chemistry

Fungal Proteins: CL, classification

*Galactose Oxidase: CH, chemistry

Galactose Oxidase: CL, classification

Molecular Sequence Data

Protein Folding

Protein Structure, Tertiary

Sequence Alignment

Sequence Homology, Amino Acid

CN EC 1.1.3.9 (Galactose Oxidase); 0 (discoidins); 0 (Bacterial Proteins); 0 (Fungal Proteins)

L157 ANSWER 10 OF 23 MEDLINE

AN 1996309401 MEDLINE

DN 96309401

TI Phylogenetic relationships of fungi, plantae, and animalia inferred from homologous comparison of ribosomal proteins.

AU Veuthey A L; Bittar G

CS Departement de Biochimie Medicale, Universite de Genève, Switzerland.

L157 AM2WFR 11 OF 23 MEDLINE
AN 1998090839 MEDLINE
DN 27.5 V38
TI The expanding beta 4-galactosyltransferase gene family: messages from the databanks.
AU Lo N W; Shaper J H; Pevsner J; Shaper N L
CE Department of Pharmacology and Molecular Sciences, Kennedy Krieger Institute, Baltimore, MD, USA.
NC CA45799 (NCI)
SO GLYCOPROTEINS, (1998 May) 8 (5) 517-26.
Journal code: BEL. ISSN: 0889-6593.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
OS GENBANK-AF038660; GENBANK-AF038661; GENBANK-AF038662; GENBANK-AF038663; GENBANK-AF038664
EM 199810
AB From a systematic search of the UniGene and dbEST databanks, using human beta 4-galactosyltransferase (beta 4GalT-I), which is recognized to function in lactose biosynthesis, as the query sequence, we have identified five additional gene family members denoted as beta 4GalT-II, -III, -IV, -V, and -VI. Complementary DNA clones containing the complete coding regions for each of the five human homologs were obtained or generated by a PCR-based strategy (RACE) and sequenced. Relative to beta 4GalT-I, the percent sequence identity at the amino acid level between the individual family members, ranges from 33% (beta 4GalT-VI) to 56% (beta 4GalT-II). The highest sequence identity between any of the homologs is between beta 4GalT-V and beta 4GalT-VI (61%). beta 4GalT-II is the ortholog of the chicken beta 4GalT-II gene, which has been demonstrated to encode an alpha-lactalbumin responsive beta 4-galactosyltransferase (Shaper et al., J. Biol. Chem., 272, 31359-31399, 1997). As established by Northern analysis, beta 4GalT-II and -IV show the most restricted pattern of tissue expression. High steady state levels of beta 4GalT-II mRNA are seen only in fetal brain and adult heart, muscle, and pancreas; relatively high levels of beta 4GalT-VI mRNA are seen only in adult brain. When the corresponding mouse EST clone for each of the beta 4GalT family members

was used as the hybridization probe for Northern analysis of murine mammary tissue, transcription of only the beta 4GALT-1 gene could be detected in the lactating mammary gland. These observations support the conclusion that among the six known beta 4GALT family members in the mammalian genome, that have been generated through multiple gene duplication events of an ancestral gene(s), only the beta 4GALT-1 ancestral lineage was recruited for lactose biosynthesis during the evolution of mammals.

CT Check Tags: Animal; Human; Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.

Adult

Amino Acid Sequence

***Databases, Factual
Evolution, Molecular**

Fetus

Galactosyltransferases: BI, biosynthesis

*Galactosyltransferases: CH, chemistry

Mice

Molecular Sequence Data

*Multigene Family

Phylogeny

Polymerase Chain Reaction

Recombinant Proteins: BI, biosynthesis

Recombinant Proteins: CH, chemistry

RNA, Messenger: ME, metabolism

Sequence Alignment

Sequence Homology, Amino Acid

Transcription, Genomic

CN EC 3.4.1.- (Galactosyltransferases ; C (Recombinant Proteins); C (RNA, Messenger).

L157 ANSWER 12 OF 23 MEDLINE

AN 1098304936 MEDLINE

DN 38104936

TI Chimeric structure of the NAD(P)+- and NADP+-dependent malic enzymes of Rhizobium (*Sinorhizobium*) meliloti.

AU Mitsch M J; Viegas R T; Cowie A; Osteras M; Firsov T M

CS Department of Biology, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada.

SO JOURNAL OF BIOLOGICAL CHEMISTRY, (1998 Apr 10) 273 (15) 9310-6.

Journal code: HIV. ISSN: 0021-9258.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals; Cancer Journals

EM 109807

AB Malic enzymes catalyze the oxidative decarboxylation of malate to pyruvate in conjunction with the reduction of a nicotinamide cofactor. We determined the DNA sequence and transcriptional start sites of the genes encoding the diphosphopyridine nucleotide-dependent malic enzyme (DME, EC 1.1.1.39) and the triphosphopyridine nucleotide-dependent malic enzyme (TME, EC 1.1.1. 40) of Rhizobium (*Sinorhizobium*) meliloti. The predicted DME and TME proteins contain 770 and 764 amino acids, respectively, and are approximately 320 amino acids larger than previously characterized prokaryotic malic enzymes. The increased size of DME and TME resides in the C-terminal extensions which are similar in sequence to phosphotransacetylase enzymes (EC 2.3.1.8). Modified DME and TME proteins which lack this C-terminal region retain malic enzyme activity but are unable to oligomerize into the native state. Data base searches have revealed that similar chimeric malic enzymes were uniquely present in Gram-negative bacteria. Thus DME and TME appear to be members of a new class of malic enzyme characterized by the presence of a phosphotransacetylase-like domain at the C terminus of the protein.

CT Check Tags: Comparative Study; Support, Non-U.S. Gov't

Amino Acid Sequence

Bacteria: EN, enzymology

Bacteria: GE, genetics

Chimeric Proteins: BI, biosynthesis

Chimeric Proteins: CH, chemistry

Chimeric Proteins: ME, metabolism

***Evolution, Molecular**

*Malate Dehydrogenase: CH, chemistry

*Malate Dehydrogenase: GE, genetics

Malate Dehydrogenase: ME, metabolism

Molecular Sequence Data

Phylogeny

Sequence Alignment

Sequence Homology, Amino Acid

**Chlorobaculum meliliti*: EN, enzymology

Chlorobaculum meliliti: GE, genetics

Software

CN EC 1.1.1.37 (Malate Dehydrogenase); EC 1.1.1.39 (malate dehydrogenase (decarboxylating)); EC 1.1.1.40 (malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+)); 0 (Chimeric Proteins)

L157 ANSWER 13 OF 23 MEDLINE

AN 1093181064 MEDLINE

DN 3a151064

TI Oligoconsensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences.

AU Rose T M; Schultz E R; Henikoff S G; Pietrokowski S; McCallum C M; Henikoff S

CS Department of Pathobiology, School of Public Health and Community Medicine, University of Washington, Seattle, WA 98195, USA.

SO NUCLEIC ACIDS RESEARCH, (1998 Apr 1) 26 (7) 1622-35.

Journal code: OBL. ISSN: 0305-1048.

CY ENGLAND: United Kingdom

CT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals; Cancer Journals

EM 199807

EW 19980705

AB We describe a new primer design strategy for PCR amplification of unknown targets that are related to multiply-aligned protein sequences. Each primer consists of a short 3' degenerate core region and a longer 5' consensus clamp region. Only 3-4 highly conserved amino acid residues are necessary for design of the core, which is stabilized by the clamp during annealing to template molecules. During later rounds of amplification, the non-degenerate clamp permits stable annealing to product molecules. We demonstrate the practical utility of this hybrid primer method by detection of diverse reverse transcriptase-like genes in a human genome, and by detection of CEDNA methyltransferase homologs in various plant DNAs. In each case, amplified products were sufficiently pure to be cloned without gel fractionation. This Oligoconsensus-DEgenerate Hybrid Oligonucleotide Primer (CODEHOP) strategy has been implemented as a computer program that is accessible over the World Wide Web (<http://blocks.fhcrc.org/codehop.html>) and is directly linked from the BlockMaker multiple sequence alignment site for hybrid primer prediction beginning with a set of related protein sequences.

CT Check Tags: Animal; Human; Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.

Amino Acid Sequence

Arthritis, Rheumatoid: GE, genetics

Base Sequence

Codon

Computer Communication Networks

Consensus Sequence

Conserved Sequence

*DNA Modification Methylases: CH, chemistry

DNA Modification Methylases: GE, genetics

*DNA Primers

***Evolution, Molecular**

Molecular Sequence Data
 Nucleic Acid Hybridization
***Phylogeny**
 Polymerase Chain Reaction: MT, methods
***RNA-Directed DNA Polymerase: CH, chemistry**
RNA-Directed DNA Polymerase: GE, genetics
 Sarcoma, Kaposi: GE, genetics
Sequence Alignment
Sequence Homology, Amino Acid
 Sequence Homology, Nucleic Acid
Software

CN EC 3.1.1.1 (DNA Modification Methylases); EC 2.7.7.43 (RNA-Directed DNA Polymerase); 3 (Codon); 6 (DNA Primers)

L157 ANSWER 14 OF 23 MEDLINE

AN 1998169454 MEDLINE

DN 99169454

TI Seeking an ancient enzyme in *Methanococcus jannaschii* using ORF, a program based on predicted secondary structure comparisons.

AU Aurora R; Rose G P

CS Department of Biophysics and Biophysical Chemistry, Johns Hopkins School of Medicine, 720 North Wolfe Street, Baltimore, MD 21205, USA.

SO PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, (1998 Mar 17) 95 (6) 2818-23.

Journal code: PNAS. ISSN: 0027-8424.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals; Cancer Journals

EM 1998-99

EW 1998-99

AB We have developed a simple procedure to identify protein homologs in genomic databases. The program, called ORF, is based on comparisons of predicted secondary structure. Protein structure is far better conserved than amino acid sequence, and structure-based methods have been effective in exploiting this fact to find homologs, even among proteins with scant sequence identity. ORF is a secondary structure-based method that operates solely on predictions from sequence and requires no experimentally determined information about the structure. The approach is illustrated by an example: Thymidylate synthase, a highly conserved enzyme essential to thymidine biosynthesis in both prokaryotes and eukaryotes, is thought to be used by Archaea, but a corresponding gene has yet to be identified. Here, a candidate thymidylate synthase is identified as a previously unassigned open reading frame from the genome of *Methanococcus jannaschii*, viz., MJ0757. Using primary structure information alone, the optimally aligned sequence identity between MJ0757 and *Escherichia coli* thymidylate synthase is 7%, well below the threshold of sensitivity for detection by sequence-based methods.

CT Check Tags: Comparative Study; Support, U.S. Gov't, P.H.S.
 Amino Acid Sequence

Archaeal Proteins: CH, chemistry

Bacteria: EN, enzymology

Evolution, Molecular

**Methanococcus*: EN, enzymology

Models, Molecular

Molecular Sequence Data

*Protein Structure, Secondary

Protein Structure, Tertiary

Saccharomyces cerevisiae: EN, enzymology

***Sequence Alignment: MT, methods**

Sequence Homology, Amino Acid

***Software**

*Thymidylate Synthase: CH, chemistry

CN EC 3.1.1.43 (Thymidylate Synthase); 0 (Archaeal Proteins)

L157 ANSWER 15 OF 23 MEDLINE

AN 1998051972 MEDLINE
 DN 9805197.
 TI An approach to detection of protein structural motifs using an encoding scheme of backbone conformations.
 AU Matsuda H; Taniguchi F; Hashimoto A
 CC Department of Informatics and Mathematical Science, Graduate School of Engineering Science, Osaka University, Japan.. matsuda@ics.es.osaka-u.ac.jp
 SO PACIFIC SYMPOSIUM ON BIocomputing, (1997) 280-91.
 CY Singapore
 DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 PR Priority Journals
 EM 199803
 EW 19910101
 AB This paper presents an approach to detection of protein structural motifs. In our approach, first all protein backbone conformations are converted into character strings using an encoding scheme. Then we use the Smith-Waterman local alignment algorithm to detect common structural motifs. By comparing results with the PROSITE regular expression patterns, our method can detect several motifs which the PROSITE patterns fail to detect.
 JT Check Tags: Support, Non-U.S. Gov't
 Amino Acid Sequence
 Calcium-Binding Proteins: CH, chemistry
 Computer Simulation
 Crystallins: CH, chemistry
 Databases
 Databases, Factual
 Evolution, Molecular
 Information Storage and Retrieval
 Models, Molecular
 Molecular Sequence Data
 *Protein Conformation
 Protein Structure, Secondary
 *Proteins: CH, chemistry
 *Proteins: GE, genetics
 Sequence Alignment
 Sequence Homology, Amino Acid
 Software
 CN 0 (Calcium-Binding Proteins); 0 (Crystallins); 0 (Proteins)

L157 ANSWER 16 OF 23 MEDLINE
 AN 1998001595 MEDLINE
 DN 98001595
 TI Evolutionary conservation and predicted structure of the *Drosophila* extra sex comb repressor protein.
 AU Ng J; Li P; Morgan K; Simon J
 CC Department of Biochemistry, University of Minnesota, St. Paul 55108, USA.
 NC GM49850 (NIGMS)
 SO MOLECULAR AND CELLULAR BIOLOGY, (1997 Nov) 17 (11) 6663-72.
 Journal code: NGY. ISSN: 0270-7306.
 CY United States
 DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 PR Priority Journals
 CC GENBANK-AF003605; GENBANK-AF003604; GENBANK-AF003605
 EM 199801
 AB The *Drosophila* extra sex comb (esc) protein, a member of the Polycomb group (PcG), is a transcriptional repressor of homeotic genes. Genetic studies have shown that esc protein is required in early embryos at about the time that other PcG proteins become engaged in homeotic gene repression. The esc protein consists primarily of multiple copies of the WD repeat, a motif that has been implicated in protein-protein interaction. To further investigate the domain organization of esc

protein, we have isolated and characterized esc homologs from divergent insect species. We report that esc protein is highly conserved in housefly (71% identical to *Drosophila* esc), butterfly (55% identical), and grasshopper (54% identical). We show that the butterfly homolog provides esc function in *Drosophila*, indicating that the sequence similarities reflect functional conservation. Homology modeling using the crystal structure of another WD repeat protein, the G-protein beta-subunit, predicts that esc protein adopts a beta-propeller structure. The sequence comparisons and modeling suggest that there are seven WD repeats in esc protein which together form a seven-bladed beta-propeller. We locate the conserved regions in esc protein with respect to this predicted structure. Site-directed mutagenesis of specific loops, predicted to extend from the propeller surface, identified conserved parts of esc protein required for function *in vivo*. We suggest that these regions might mediate physical interaction with esc partner proteins.

CT Check Tags: Animal; Support, U.S. Gov't, Non-F.H.S.; Support, U.S. Gov't, F.H.S.

Amino Acid Sequence

Butterflies: GE, genetics

Cloning, Molecular

Computer Simulation

Conserved Sequence

Drosophila: EM, embryology

*Drosophila: GE, genetics

DNA, Complementary: GE, genetics

*Evolution, Molecular

Grasshoppers: GE, genetics

Houseflies: GE, genetics

Insect Proteins: CH, chemistry

*Insect Proteins: GE, genetics

Insect Proteins: ME, metabolism

Models, Molecular

Molecular Sequence Data

Mutagenesis, Site-Directed

Protein Binding

Protein Conformation

Protein Structure, Tertiary

Repressor Proteins: CH, chemistry

*Repressor Proteins: GE, genetics

Repressor Proteins: ME, metabolism

Sequence Analysis, DNA

Sequence Homology, Amino Acid

Structure-Activity Relationship

FN 134549-01-2 (Polycomb protein, insect)

CN 0 (DNA, Complementary); 0 (Insect Proteins); 0 (Repressor Proteins)

L157 ANSWER 17 OF 23 MEDLINE

AN 97435515 MEDLINE

DN 97435515

TI Homeoboxes in sea anemones (Cnidaria:Anthozoa): a PCR-based survey of *Nematostella vectensis* and *Metridium senile*.

AU Finnerty J R; Martindale M Q

CS Department of Organismal Biology and Anatomy, University of Chicago, Illinois 60637, USA.

NC HD07136 (NICHD)

SO BIOLOGICAL BULLETIN, (1997 Aug) 193 (1) 63-76.

Journal code: A26. ISSN: 0006-3185.

CY United States

CT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

OS GENBANK-U42726; GENBANK-U42727; GENBANK-U42728; GENBANK-U42729;

GENBANK-U42730; GENBANK-U42731; GENBANK-U42732; GENBANK-U42733;

GENBANK-U42734; GENBANK-U42735; GENBANK-U42736; GENBANK-U42737;

GENBANK-M62872; GENBANK-S15548; GENBANK-M62871; GENBANK-D44629;

GENBANK-D37042; GENBANK-L09690; GENBANK-E37042; GENBANK-A35511;

GENBANK-C44636; GENBANK-I44629; GENBANK-F44630; GENBANK-H48200;
 GENBANK-S39563; GENBANK-S75228; GENBANK-S36773; GENBANK-A60042;
 GENBANK-XC5138; GENBANK-S22516; +

EM 199712

EW 19971201

AB Homeobox genes belong to a phylogenetically widespread family of regulatory genes that play important roles in pattern formation and cell-fate specification in several model systems (e.g., *Drosophila*, mouse, and *C. elegans*). Although the evolution of many classes of homeobox genes predates the diversification of the Bilateria, comparatively little is known about homeobox genes in outgroups to the Bilateria, such as the Cnidaria. We used the polymerase chain reaction to recover 12 partial homeoboxes from 2 species of sea anemones, *Metridium senile* and *Nematostella vectensis* (phylum Cnidaria; class Anthozoa). These homeoboxes appear to represent 9 distinct, mutually paralogous homeobox genes, 5 of which belong to previously identified cnidarian homeobox classes, and 4 of which appear to represent previously unidentified classes. The evolutionary relationships between the homeodomains of sea anemones and of bilaterian animals were assessed through database searches and phylogenetic analyses. As many as 5 of the anemone homeoboxes may belong to the Hox class, which suggests that the Hox gene complement of cnidarians is larger than previously expected. Homologs of the even-skipped gene of *Drosophila* were also identified in both *Metridium* and *Nematostella*.

CT Click Tags: Animal; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.

Amino Acid Sequence

Base Sequence

Meenhorstadiis: GE, genetics

Cloning, Molecular

Drosophila: GE, genetics

*Evolution

Evolution, Molecular

*Genes, Homeobox

Homeodomain Proteins: BI, biosynthesis

*Homeodomain Proteins: CH, chemistry

Homeodomain Proteins: GE, genetics

Information Systems

Mice

Molecular Sequence Data

Phylogeny

Polymerase Chain Reaction: MT, methods

Recombinant Proteins: BI, biosynthesis

Recombinant Proteins: CH, chemistry

Sea Anemones: CL, classification

*Sea Anemones: GE, genetics

Sequence Alignment

Sequence Homology, Amino Acid

Sequence Homology, Nucleic Acid

CN 0 (Homeodomain Proteins); 0 (Recombinant Proteins)

L157 ANSWER 18 OF 23 MEDLINE

AN 37422608 MEDLINE

DN 37422608

TI The human beta-defensin-1 and alpha-defensins are encoded by adjacent genes: two peptide families with differing disulfide topology share a common ancestry.

AU Liu L; Zhao C; Heng H H; Ganz T

CS Department of Medicine and Will Rogers Institute for Pulmonary Research, University of California at Los Angeles School of Medicine, 90095, USA.

NC HL-07014 (NHLBI)
 HL-35640 (NHLBI)
 HL-46809 (NHLBI)

SO JENOMICS, (1997 Aug 1) 43 (3) 316-20.

CY Journal code: GEN. ISSN: 0883-7543.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

OS GENBANK-U56920; GENBANK-U50931; GENBANK-X92744

EM 19901

AB We cloned a novel human beta-defensin gene and determined its full-length cDNA sequence. The entire gene spanned more than 7 kb and included a large 662-kp intron. The 261-kp cDNA encoded a prepropeptide that corresponded precisely to the recently identified human beta-defensin HBD-1, an antimicrobial peptide implicated in the resistance of epithelial surfaces to microbial colonization. By two-color fluorescence *in situ* hybridization on both metaphase chromosome and released chromatin fiber, HBD-1 gene (HPEB1 in HUGO/GDB nomenclature) mapped to chromosomal region 6p23.1-p23.2 in close proximity (within 100-150 kb) to the gene for the human neutrophil alpha-defensin HNE-1 (DEFAL1). Thus, despite a complete lack of DNA sequence similarity and despite differences in their disulfide-pairing pattern, the alpha- and beta-families appear to have evolved from a common premammalian defensin gene.

CT Check Tags: Human; Support, U.C. Gov't, P.H.S.

Algorithms

Amino Acid Sequence

*Anti-Infective Agents: ME, metabolism

Base Sequence

*Blood Proteins: GE, genetics

Chromosome Mapping

Chromosomes, Human, Pair 8: GE, genetics

Databases, Factual

Disulfides: CH, chemistry

DNA, Complementary: AN, analysis

DNA, Complementary: CH, chemistry

DNA, Complementary: GE, genetics

Evolution, Molecular

*Genes, Structural: GE, genetics

In Situ Hybridization, Fluorescence

Molecular Sequence Data

*Multigene Family

Peptides: CH, chemistry

Peptides: GE, genetics

Protein Structure, Secondary

Sequence Alignment: MT, methods

Sequence Homology, Amino Acid

CN 0 (defensins); 0 (Anti-Infective Agents); 0 (Blood Proteins); 0 (Disulfides); 0 (DNA, Complementary); 0 (Peptides)

LI57 ANSWER 19 OF 23 MEDLINE

AN 97233150 MEDLINE

DN 97233150

TI Optimum superimposition of protein structures: ambiguities and implications.

AU Feng Z K; Sippl M J

CS Center for Applied Molecular Engineering, University of Salzburg, Australia.

SO FOLDING AND DESIGN, (1996) 1 (2) 125-32.

Journal code: CUD. ISSN: 1369-0278.

CY ENGLAND: United Kingdom

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

EM 199707

EW 19970701

AB **BACKGROUND:** Techniques for comparison and optimum superimposition of protein structures are indispensable tools, providing the basis for statistical analysis, modeling, prediction and classification of protein folds. Observed similarity of structures is frequently interpreted as an indication of evolutionary relatedness. A variety of advanced techniques are available, but so far the important issue of uniqueness of structural

superimposition has been largely neglected. We set out to investigate this issue by implementing an efficient algorithm for structure superimposition enabling routine searches for alternative alignments. RESULTS: The algorithm is based on optimum superimposition of structures and dynamic programming. The implementation is tested and validated using published results. In particular, an automatic classification of all protein folds in a recent release of the protein data bank is performed. The results obtained are closely related to published data. Surprisingly, for many protein pairs alternative alignments are obtained. These alignments are indistinguishable in terms of number of equivalent residues and root mean square error of superimposition, but the respective sets of equivalent residue pairs are completely distinct. Alternative alignments are observed for all protein architectures, including mixed alpha/beta folds.

CONCLUSIONS: Superimposition of protein folds is frequently ambiguous. This has several implications on the interpretation of structural similarity with respect to evolutionary relatedness and it restricts the range of applicability of superimposed structures in statistical analysis. In particular, studies based on the implicit assumption that optimum superimposition of structures is unique are bound to be misleading.

CT Check Tags: Support, Non-U.S. Gov't

***Algorithms**

Amino Acid Sequence
 Evaluation Studies
Evolution, Molecular
 Models, Molecular
 Molecular Sequence Data
 Molecular Structure
 Protein Folding
 Protein Structure, Secondary
***Proteins: CH, chemistry**
Proteins: GE, genetics
Sequence Alignment
Sequence Homology, Amino Acid

CN 0 (Proteins)

L157 ANSWER TO OF 2? MEDLINE

AN 97224626 MEDLINE

DN 97224626

TI Evolutionary motif and its biological and structural significance.

AU Tateno Y; Ikeo K; Imanishi T; Watanabe H; Endo T; Yamaguchi Y; Suzuki Y; Takahashi K; Tsunoyama K; Kawai M; Kawanishi Y; Naitou K; Gojobori T

CS Center for Information Biology, National Institute of Genetics, Mishima, Japan.. ytateno@genes.nig.ac.jp

SO JOURNAL OF MOLECULAR EVOLUTION, (1997) 44 Suppl 1 S38-43.

Journal code: J76. ISSN: 0022-2844.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

EM 199706

EW 19970601

AB We developed a method for multiple alignment of protein sequences. The main feature of this method is that it takes the evolutionary relationships of the proteins in question into account repeatedly for execution, until the relationships and alignment results are in agreement. We then applied this method to the data of the international DNA sequence databases, which are the most comprehensive and updated DNA databases in the world, in order to estimate the "evolutionary motif" by extensive use of a supercomputer. Though a few problems needed to be solved, we could estimate the length of the motifs in the range of 20 to 200 amino acids, with about 60 the most frequent length. We then discussed their biological and structural significance. We believe that we are now in a position to analyze DNA and protein not only in vivo and in vitro but also in silico.

CT Check Tags: Animal; Human

Amino Acid Sequence

Amino Acids: GE, genetics

Databases, Factual

DNA: GE, genetics

Evolution, Molecular**Proteins: GE, genetics**

Sequence Alignment

***Sequence Homology, Amino Acid**

RN 9607-43-2 (DNA)

CN 0 (Amino Acids); 0 (Proteins)

L157 ANSWER 11 OF 23 MEDLINE

AN 97163541 MEDLINE

DN 97163541

TI Of worms and men: an evolutionary perspective on the fibroblast growth factor (FGF) and FGF receptor families.

AU Chaulier F; Pontarotti P; Boumin R; Hartung H; Goldfarb M; Birnbaum D

CS Laboratoire d'Oncoologie Moléculaire, U119 INSERM, 27 Bd. Lez Escale, 13009 Marseille, France.

SO JOURNAL OF MOLECULAR EVOLUTION, (1997 Jan) 44 (1) 43-56.

Journal code: J76. ISSN: 0022-2844.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

GS GENBANK-AJ001055?

EM 1-49775

AB FGFs (fibroblast growth factors) play major roles in a number of developmental processes. Recent studies of several human disorders, and concurrent analysis of gene knock-out and properties of the corresponding recombinant proteins have shown that FGFs and their receptors are prominently involved in the development of the skeletal system in mammals. We have compared the sequences of the nine known mammalian FGFs, FGFs from other vertebrates, and three additional sequences that we extracted from existing databases: two human FGF sequences that we tentatively designated FGF10 and FGF11, and an FGF sequence from *Caenorhabditis elegans*. Similarly, we have compared the sequences of the four FGF receptor paralogous found in chordates with four non-chordate FGF receptors, including one recently identified in *C. elegans*. The comparison of FGF and FGF receptor sequences in vertebrates and nonvertebrates shows that the FGF and FGF receptor families have evolved through phases of gene duplications, one of which may have coincided with the emergence of vertebrates, in relation with their new system of body scaffold.

CT Check Tags: Animal; Human; Support, Non-U.S. Gov't

Amino Acid Sequence

Caenorhabditis elegans: GE, genetics**Databases, Factual*****Evolution, Molecular*****Fibroblast Growth Factor: GE, genetics**

Genes, Structural

Molecular Sequence Data

Multigene Family

Phylogeny

***Receptors, Fibroblast Growth Factor: GE, genetics**

Sequence Alignment

Sequence Homology, Amino Acid

***Vertebrates: GE, genetics**

RN 62031-54-3 (Fibroblast Growth Factor)

CN 0 (Receptors, Fibroblast Growth Factor)

L157 ANSWER 12 OF 23 MEDLINE

AN 96404124 MEDLINE

DN 96404124

TI Teleost HoxD and HoxA genes: comparison with tetrapods and functional evolution of the HOXD complex.

AU van der Hoeven F; Sordino P; Fraudeau N; Izpisua-Belmonte J C; Duboule D

CS Department of Zoology and Animal Biology, University of Geneva, Switzerland.

SO MECHANISMS OF DEVELOPMENT, (1996 Jan) 54 (1) 9-21.
 Journal code: AXF. ISSN: 0925-4773.
 CY Ireland
 DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 FS Priority Journals
 DS GENBANK-X87750; GENBANK-X87751; GENBANK-X87752; GENBANK-X87753
 EM 199702
 AB In tetrapods, Hox genes are essential for the proper organization and development of axial structures. Experiments involving Hox gene inactivations have revealed their particularly important functions in the establishment of morphological transitions within metameric series such as the vertebral column. Teleost fish show a much simpler range of axial (trunk or appendicular) morphologies, which prompted us to investigate the nature of the Hox system in these lower vertebrates. Here, we show that fish have a family of Hox genes, very similar in both number and general organization, to that of tetrapods. Expression studies, carried out with HoxE and HoxA genes, showed that all vertebrates use the same general scheme, involving the colinear activation of gene expression in both space and time. Comparisons between tetrapods and fish allowed us to propose a model which accounts for the primary function of this gene family. In this model, a few ancestral Hox genes were involved in the determination of polarity in the digestive tract and were further recruited in more elaborate axial structures.
 CT Check Tags: Animal; Comparative Study; Support, Non-U.S. Gov't
 Amino Acid Sequence
 Chick Embryo
 Digestive System: EM, embryology
 Digestive System: ME, metabolism
 DNA, Complementary: GE, genetics
 *Evolution, Molecular
 Fetal Development: GE, genetics
 Gene Expression Regulation, Developmental
 *Genes, Homeobox
 Homeodomain Proteins: BI, biosynthesis
 *Homeodomain Proteins: GE, genetics
 In Situ Hybridization
 Mice
 Models, Biological
 Molecular Sequence Data
 Morphogenesis: GE, genetics
 RNA Probes
 Sequence Alignment
 Sequence Homology, Amino Acid
 Species Specificity
 Trans-Activators: BI, biosynthesis
 *Trans-Activators: GE, genetics
 Transcription Factors: BI, biosynthesis
 *Transcription Factors: GE, genetics
 Urogenital System: EM, embryology
 Urogenital System: ME, metabolism
 Vertebrates: CL, classification
 Vertebrates: EM, embryology
 *Vertebrates: GE, genetics
 Zebrafish: EM, embryology
 *Zebrafish: GE, genetics
 RN 157907-48-7 (HoxA protein)
 CN 0 (DNA, Complementary); 0 (Homeodomain Proteins); 0 (Hoxd-13 protein); 0 (HoxD-10 protein); 0 (HoxD-11 protein); 0 (HoxD-12 protein); 0 (RNA Probes); 0 (Trans-Activators); 0 (Transcription Factors)
 L157 ANSWER 23 OF 23 MEDLINE
 LN 96180040 MEDLINE
 DN 96180020
 TI An evolutionary trace method defines binding surfaces common to protein families.

AU Lichtarge O; Bourne H R; Cohen F E
 CS Department of Cellular and Molecular Pharmacology, University of
 California San Francisco, 94143-0450, USA.
 SO JOURNAL OF MOLECULAR BIOLOGY, (1996 Mar 29) 267 (2) 342-58.
 Journal code: JMB. ISSN: 0022-2836.
 CY ENGLAND: United Kingdom
 DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 FS Priority Journals; Cancer Journals
 SM 199608
 AB X-ray or NMR structures of proteins are often derived without their
 ligands, and even when the structure of a full complex is available, the
 area of contact that is functionally and energetically significant may be
 a specialized subset of the geometric interface deduced from the spatial
 proximity between ligands. Thus, even after a structure is solved, it
 remains a major theoretical and experimental goal to localize protein
 functional interfaces and understand the role of their constituent
 residues. The evolutionary trace method is a systematic, transparent and
 novel predictive technique that identifies active sites and functional
 interfaces in proteins with known structure. It is based on the extraction
 of functionally important residues from sequence conservation patterns in
 homologous proteins, and on their mapping onto the protein surface to
 generate clusters identifying functional interfaces. The SH2 and SH3
 modular signaling domains and the DNA binding domain of the nuclear
 hormone receptors provide tests for the accuracy and validity of our
 method. In each case, the evolutionary trace delineates the functional
 epitope and identifies residues critical to binding specificity. Based on
 mutational evolutionary analysis and on the structural homology of protein
 families, this simple and versatile approach should help focus
 site-directed mutagenesis studies of structure-function relationships in
 macromolecules, as well as studies of specificity in molecular
 recognition. More generally, it provides an evolutionary perspective for
 judging the functional or structural role of each residue in protein
 structure.
 CT Check Tags: Animal; Human; Support, Non-U.S. Gov't; Support, U.S. Gov't,
 P.H.S.
 src Homology Domains
 Amino Acid Sequence
 *Binding Sites
 Conserved Sequence
 Databases, Factual
 DNA-Binding Proteins: CH, chemistry
 *Evolution, Molecular
 Models, Molecular
 Molecular Sequence Data
 Mutation
 Protein Conformation
 *Proteins: CH, chemistry
 Rats
 Receptors, Glucocorticoid: CH, chemistry
 Sequence Alignment: MT, methods
 Sequence Homology, Amino Acid
 Sinc: Fingers
 CN 0 (DNA-Binding Proteins); 0 (Proteins); 0 (Receptors, Glucocorticoid)

=> d ail tot

L162 ANSWER 1 OF 8 MEDLINE
 AN 1998452316 MEDLINE
 DN 98452316
 TI Sequence, structural, functional, and phylogenetic analyses of three
 glycosidase families.
 AU Mian I S
 CS Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley,

CA 94/20, USA.. SMian@lbl.gov
SO BLOOD CELLS, MOLECULES, AND DISEASES, (1993 Jun) 24 (2) 83-100.
Journal code: BSA. ISSN: 1078-9796.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199301
EW 19930104
AB Glycosidases, which cleave the glycosidic bond between a carbohydrate and another moiety, have been classified into over 63 families. Here, a variety of computational techniques have been employed to examine three families important in normal and abnormal pathology with the aim of developing a framework for future homology modeling, experimental and other studies. Family 1 includes bacterial and archaeal enzymes as well as lactase-phlorizin-hydrolase and *lacto-N*-glycosidase implicated in disaccharide intolerance II and aging respectively. A statistical model, a hidden Markov model (HMM), for the family 1 glycosidase domain was trained and used as the basis for comparative examination of the conserved and variable sequence and structural features as well as the phylogenetic relationships between family members. Although the structures of four family 1 glycosidases have been determined, this is the first comparative examination of all these enzymes. Aspects that are unique to specific members or subfamilies (substrate binding loops) as well those common to all members (a beta/alpha)8 barrel fold) have been defined. Active site residues in some domains in *lacto-N* and lactase-phlorizin hydrolases differ from other members and in one instance may bind but not cleave substrate. The four invariant and most highly conserved residues are not residues implicated in catalysis and/or substrate binding. Of these, a histidine may be involved in transition state stabilization. Glucosylceramidase (family 30) and galactosylceramidase (family 59) are mutated in the lysosomal storage disorders Gaucher disease and Krabbe disease, respectively. HMM-based analysis, structure prediction studies and examination of disease mutations reveal a glycosidase domain common to these two families that also occurs in some bacterial glycosidases. Similarities in the reactions catalyzed by families 30 and 59 are reflected in the presence of a structurally and functionally related (beta/alpha)8 barrel fold related to that in family 1.
CT Check Tags: Animal; Comparative Study; Human; Support, U.S. Gov't, Non-P.H.S.
Amino Acid Sequence
Amino Acid Substitution
Bacteria: EN, enzymology
Bacterial Proteins: CH, chemistry
Bacterial Proteins: CL, classification
Bacterial Proteins: GE, genetics
Binding Sites
Evolution, Molecular
Glycoside Hydrolases: CH, chemistry
*Glycoside Hydrolases: CL, classification
Glycoside Hydrolases: GE, genetics
Invertebrates: CH, chemistry
Invertebrates: CL, classification
Invertebrates: GE, genetics
Mammals: ME, metabolism
Markov Chains
Models, Molecular
Molecular Sequence Data
Multigene Family
Phylogeny
Plant Proteins: CH, chemistry
Plant Proteins: CL, classification
Plant Proteins: GE, genetics
Plants: EN, enzymology
*Protein Conformation
Sequence Alignment

Sequence Homology, Amino Acid
Structure-Activity Relationship

CN EC 3.2.1. (Glycoside Hydrolases); 0 (Bacterial Proteins); 0 (Plant
Proteins)

L162 ANSWER 2 OF 3 MEDLINE
AN 1998417653 MEDLINE
DN 99417653
TI Supersites within superfolds. Binding site similarity in the absence of
homology.
AU Russell R B; Jasieni P D; Sternberg M J E
JC Biomolecular Modelling Laboratory, Lincoln's Inn Fields, 80 Box 123,
London, WC2A 3PK, UK.
SO JOURNAL OF MOLECULAR BIOLOGY, (1998 Oct 2) 282 (4) 903-18.
Journal code: JMB. ISSN: 0022-2833.
CY ENGLAND: United Kingdom
OT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals; Cancer Journals
EM 199809
EW 19980903
AB A method is presented to assess the significance of binding site
similarities within superimposed protein three-dimensional (3D) structures
and applied to all similar structures in the Protein Data Bank. For
similarities between 3D structures lacking significant sequence
similarity, the important distinction was made between remote homology (an
ancient common ancestor) and analogy (likely convergence to a folding
motif) according to the structural classification of proteins (SCOP)
database. Supersites were defined as structural locations or groups of
analogous proteins (i.e. superfolds) showing a statistically significant
tendency to bind substrates despite little evidence of a common ancestor
for the proteins considered. We identify three potentially new superfolds
containing supersites: ferredoxin-like folds, four-helical bundles and
double-stranded beta helices. In addition, the method quantifies binding
site similarities within homologous proteins and previously identified
supersites such as that found in the beta/alpha (TIM) barrels. For the
nine superfolds, the accuracy of predictions of binding site locations is
assessed. Implications for protein evolution, and the prediction of
protein function either through fold recognition or tertiary structure
comparison, are discussed. Copyright 1998 Academic Press.
CT Check Tags: Animal; Human; Support, Non-U.S. Gov't
*Binding Sites
*Databases, Factual
Evolution, Molecular
Ferredoxins: ME, metabolism
Ligands
Models, Molecular
Protein Conformation
*Protein Folding
Protein Structure, Secondary
*Proteins: CH, chemistry
Proteins: ME, metabolism
Sequence Alignment
Sequence Homology, Amino Acid
Statistical Distributions
Structure-Activity Relationship
CN 0 (Ferredoxins); 0 (Ligands); 0 (Proteins)

L162 ANSWER 3 OF 3 MEDLINE
AN 1998362531 MEDLINE
DN 98362531
TI Protein disorder and the evolution of molecular recognition: theory,
predictions and observations.
AU Dunker A K; Garner E; Guillot S; Romero P; Albrecht K; Hart J; Obradovic
Z; Kissinger C; Villafranca J E
CS Department of Biochemistry & Biophysics, Washington State University,

SO Pullman 99164-4660, USA.. dunker@mail.wsu.edu
 PACIFIC SYMPOSIUM ON BIocomputing, (1998) 473-84.
 Journal code: CWO.

CY Singapore
 DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 FS Priority Journals
 EM 199901

AB Observations going back more than 20 years show that regions in proteins with disordered backbones can play roles in their binding to other molecules; typically, the disordered regions become ordered upon complex formation. Thought-experiments with Schulz diagrams, which are defined herein, suggest that disorder-to-order transitions are required for natural selection to operate separately on affinity and specificity. Separation of affinity and specificity may be essential for fine-tuning the molecular interaction networks that comprise the living state. For low affinity, high specificity interactions, our analysis suggests that natural selection would parse the amino acids conferring flexibility in the unbound state from those conferring specificity in the bound state. For high affinity, low specificity or for high affinity, multiple specificity interactions, our analysis suggests that the disorder-to-order transitions enable alternative packing interactions between side chains to accommodate the different binding targets. Disorder-to-order transitions upon binding also have significant kinetic implications as well, by having complex effects on both on- and off-rates. Current data are insufficient to decide on these proposals, but sequence and structure analysis on two examples support further investigations of the role of disorder-to-order transitions upon binding.

CT Algorithms
 Amino Acid Sequence
 Binding Sites
 Calmodulin: CH, chemistry
 *Evolution, Molecular
 Macromolecular Systems
 *Protein Conformation
 Protein Structure, Secondary
 *Proteins: CH, chemistry
 Proteins: GE, genetics
 Selection (Genetics)
 Sequence Alignment
 Sequence Homology, Amino Acid
 *Software

CH 0 (Calmodulin); 0 (Macromolecular Systems); 0 (Proteins)

L162 ANSWER 4 OF 2 MEDLINE
 AN 1998330762 MEDLINE
 DN 34330762
 TI Protein fold irregularities that hinder sequence analysis.
 AU Russell F B; Ponting C P
 CC SmithKline Beecham Pharmaceuticals, Bioinformatics, New Frontiers Science Park (North), Essex, UK.. russellr1@mh.uk.sphrd.com
 SO CURRENT OPINION IN STRUCTURAL BIOLOGY, (1998 Jun) 8 (3) 364-71. Ref: 70
 Journal code: C9V. ISSN: 0959-440X.
 CY ENGLAND: United Kingdom
 DT Journal; Article; (JOURNAL ARTICLE)
 General Review; (REVIEW)
 (REVIEW, TUTORIAL)
 LA English
 FS Priority Journals
 EM 199812
 EW 19981201
 AB The detection of homologous protein sequences frequently provides useful predictions of function and structure. Methods for homology searching have continued to improve, such that very distant evolutionary relationships can now be detected. Little attention has been paid, however, to the problems of detecting homology when domains are inserted or permuted. Here

we review recent occurrences of these phenomena and discuss methods that permit their detection.

CT Check Tags: Animal; Human; Support, Non-U.S. Gov't
 Amino Acid Sequence
 Evolution, Molecular
 Models, Molecular
 Molecular Sequence Data
 *Protein Folding
 *Protein Structure, Secondary
 *Proteins: CH, chemistry
 *Proteins: ME, metabolism
 Sequence Alignment
 Sequence Homology, Amino Acid
 CN 0 *Proteins

LI62 ANSWER 9 OF 3 MEDLINE

AN 1998263252 MEDLINE

DN 98263252

TI Protein folding and protein evolution: common folding nucleus in different subfamilies of c-type cytochromes?

AU Ptitsyn O B

CS Laboratory of Experimental and Computational Biology National Cancer Institute, National Institutes of Health, Bethesda, MD 20892-5677, USA.

SO JOURNAL OF MOLECULAR BIOLOGY, (1998 May 8) 278 (3) 655-66.
 Journal code: JMB. ISSN: 0022-2836.

CY ENGLAND: United Kingdom

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals; Cancer Journals

EM 1998

EW 1998/5/8

AB Amino acid sequences of seven subfamilies of cytochromes c (mitochondrial cytochromes c, c1; chloroplast cytochromes c6, cf; bacterial cytochromes c1, c550, c551; in total 164 sequences) have been compared. Despite extensive homology within eukaryotic subfamilies, homology between different subfamilies is very weak. Other than the three heme-binding residues (Cys13, Cys14, His16, in numeration of horse cytochrome c) there are only four positions which are conserved in all subfamilies: Gly/Ala6, Phe/Tyr10, Leu/Val/Phe94 and Tyr/Trp/Phe37. In all 17 cytochromes c with known 3D-structures, these residues form a network of conserved contacts (6-94, 6-97, 10-94, 10-97 and 94-97). Especially strong is the contact between aromatic groups in positions 10 and 97, which corresponds to 13 interatomic contacts. As residues 6, 10 and residues 94, 97 are in (i, i+4) and (i, i+3) positions in the N and C-terminal helices, respectively, the above mentioned system of conserved contacts consists mainly of contacts between one turn of N-terminal helix and one turn of C-terminal helix. The importance of the contacts between interfaces of these helices has been confirmed by the existence of these contacts in both equilibrium and kinetic molten globule-like folding intermediates, as well as by mutational evidence that these contacts are involved in tight packing between the N and C-helices. Since these four residues are not involved in heme binding and have no other apparent functional role, their conservation in highly diverged cytochromes c suggests that they are of a critical importance for protein folding. The author assumes that they are involved in a common folding nucleus of all subfamilies of c-type cytochromes. Copyright 1998 Academic Press Limited.

CT Check Tags: Animal
 Amino Acid Sequence
 Bacteria: ME, metabolism
 Binding Sites
 Chloroplasts: ME, metabolism
 Conserved Sequence
 *Cytochrome c: CH, chemistry
 *Cytochrome c: GE, genetics
 *Cytochrome c: ME, metabolism
 *Evolution, Molecular

Heme: ME, metabolism
 Horses
 Kinetics
 Mitochondria: ME, metabolism
 Models, Molecular
 Molecular Sequence Data
 *Protein Conformation
 *Protein Folding
Saccharomyces cerevisiae: ME, metabolism
 Sequence Alignment
 Sequence Homology, Amino Acid
 PN 14170-98-3 (Heme); 1407-45-1 (Cytochrome c)

 L162 ANSWER 6 OF 8 MEDLINE
 AN 1998051972 MEDLINE
 DN 97051972
 TI An approach to detection of protein structural motifs using an encoding scheme of backbone conformations.
 AU Matsuda H; Taniguchi F; Hashimoto A
 CS Department of Informatics and Mathematical Science, Graduate School of Engineering Science, Osaka University, Japan.. matsuda@ics.es.osaka-u.ac.jp
 SO PACIFIC SYMPOSIUM ON BIocomPUTING, (1997) 280-91.
 Journal code: CWP.
 CY Singapore
 DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 FS Priority Journals
 EM 1998051972
 FW 1998051972
 AB This paper presents an approach to detection of protein structural motifs. In our approach, first all protein backbone conformations are converted into character strings using an encoding scheme. Then we use the Smith-Waterman local alignment algorithm to detect common structural motifs. By comparing results with the PROSITE regular expression patterns, our method can detect several motifs which the PROSITE patterns fail to detect.
 CT Check Tags: Support, Non-U.S. Gov't
 Amino Acid Sequence
 Calcium-Binding Proteins: CH, chemistry
 Computer Simulation
 Crystallins: CH, chemistry
 Databases
 Databases, Factual
 Evolution, Molecular
 Information Storage and Retrieval
 Models, Molecular
 Molecular Sequence Data
 *Protein Conformation
 Protein Structure, Secondary
 *Proteins: CH, chemistry
 *Proteins: GE, genetics
 Sequence Alignment
 Sequence Homology, Amino Acid
 Software
 PN 0 (Calcium-Binding Proteins); 0 (Crystallins); 0 (Proteins)

 L162 ANSWER 7 OF 8 MEDLINE
 AN 97233150 MEDLINE
 DN 97233150
 TI Optimum superimposition of protein structures: ambiguities and implications.
 AU Fung Z K; Sippl M J
 CS Center for Applied Molecular Engineering, University of Salzburg, Australia.
 SO FOLDING AND DESIGN, (1996) 1 (2) 123-32.

Journal code: CUD. ISSN: 1359-0278.
 CY ENGLAND: United Kingdom
 DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 FS Priority Journals
 EM 199707
 EW 19970701
 AB **BACKGROUND:** Techniques for comparison and optimum superimposition of protein structures are indispensable tools, providing the basis for statistical analysis, modeling, prediction and classification of protein folds. Observed similarity of structures is frequently interpreted as an indication of evolutionary relatedness. A variety of advanced techniques are available, but so far the important issue of uniqueness of structural superimposition has been largely neglected. We set out to investigate this issue by implementing an efficient algorithm for structure superimposition enabling routine searches for alternative alignments. **RESULTS:** The algorithm is based on optimum superimposition of structures and dynamic programming. The implementation is tested and validated using published results. In particular, an automatic classification of all protein folds in a recent release of the protein data bank is performed. The results obtained are closely related to published data. Surprisingly, for many protein pairs alternative alignments are obtained. These alignments are indistinguishable in terms of number of equivalent residues and root mean square error of superimposition, but the respective sets of equivalent residue pairs are completely distinct. Alternative alignments are observed for all protein architectures, including mixed alpha/beta folds.
CONCLUSIONS: Superimposition of protein folds is frequently ambiguous. This has several implications on the interpretation of structural similarity with respect to evolutionary relatedness and it restricts the range of applicability of superimposed structures in statistical analysis. In particular, studies based on the implicit assumption that optimum superimposition of structures is unique are bound to be misleading.
 CT Check Tags: Support, Non-U.S. Gov't
 *Algorithms
 Amino Acid Sequence
 Evaluation Studies
 Evolution, Molecular
 Models, Molecular
 Molecular Sequence Data
 Molecular Structure
 Protein Folding
 Protein Structure, Secondary
 *Proteins: CH, chemistry
 Proteins: GE, genetics
 Sequence Alignment
 Sequence Homology, Amino Acid
 CN 0 (Proteins)

L162 ANSWER 3 OF 3 MEDLINE
 AN 96180020 MEDLINE
 DN 96180020
 TI An evolutionary trace method defines binding surfaces common to protein families.
 AU Lichtarge O; Bourne H R; Cohen F E
 CS Department of Cellular and Molecular Pharmacology, University of California San Francisco, 94143-0450, USA.
 SO JOURNAL OF MOLECULAR BIOLOGY, (1996 Mar 29) 257 (2) 342-58.
 Journal code: JMB. ISSN: 0022-2836.
 CY ENGLAND: United Kingdom
 DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 FS Priority Journals; Cancer Journals
 EM 199603
 AB X-ray or NMR structures of proteins are often derived without their ligands, and even when the structure of a full complex is available, the area of contact that is functionally and energetically significant may be

a specialized subset of the geometric interface deduced from the spatial proximity between ligands. Thus, even after a structure is solved, it remains a major theoretical and experimental goal to localize protein functional interfaces and understand the role of their constituent residues. The evolutionary trace method is a systematic, transparent and novel predictive technique that identifies active sites and functional interfaces in proteins with known structure. It is based on the extraction of functionally important residues from sequence conservation patterns in homologous proteins, and on their mapping onto the protein surface to generate clusters identifying functional interfaces. The SH2 and SH3 modular signaling domains and the DNA binding domain of the nuclear hormone receptors provide tests for the accuracy and validity of our method. In each case, the evolutionary trace delineates the functional epitope and identifies residues critical to binding specificity. Based on mutational evolutionary analysis and on the structural homology of protein families, this simple and versatile approach should help focus site-directed mutagenesis studies of structure-function relationships in macromolecules, as well as studies of specificity in molecular recognition. More generally, it provides an evolutionary perspective for judging the functional or structural role of each residue in protein structure.

CT Check Tags: Animal; Human; Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.
 *src Homology Domains
 Amino Acid Sequence
 *Binding Sites
 Conserved Sequence
 Databases, Factual
 DNA-Binding Proteins: CH, chemistry
 *Evolution, Molecular
 Models, Molecular
 Molecular Sequence Data
 Mutation
 Protein Conformation
 *Proteins: CH, chemistry
 Pnts
 Receptors, Glucocorticoid: CH, chemistry
 Sequence Alignment: MT, methods
 Sequence Homology, Amino Acid
 Zinc Fingers

CN C (DNA-Binding Proteins); C (Proteins); C (Receptors, Glucocorticoid)

=> d his

(FILE 'HOME' ENTERED AT 12:33:16 ON 12 FEB 2001)
 SET COST OFF

FILE 'MEDLINE' ENTERED AT 12:34:00 ON 12 FEB 2001
 E MARCOTTE E/AU

LI 17 S E3,E4
 E EISENBERG D/AU
 LT 227 S E3-E9
 E THOMPSON M/AU
 L4 511 S E1,F14
 E PELLEGRINI M/AU
 L4 2004 S E1-E10
 E YEATES T/AU
 L5 42 S E4
 L6 975 S L1-L5
 E SEQUENCE ALIGNMENT/CT
 E E3+ALL/CT
 LT 24002 S E4+NT/CT
 L8 26 S L6 AND L7
 L9 136 S L1./CT AND L6

L10 63 S D12..,CT AND L4
E SEQUENCE ANALYSIS+ALL/CT
L11 38171 S E4+NT/CT
L12 11 S L11 AND L6
L13 3 S L10 AND L12
E SEQUENCE HOMOLOGY+ALL/CT
L14 4513-0 S E14+NT/CT
L15 354 S L6 AND L14
L16 45 S L15 AND L17
L17 11 S L8 AND L16
L18 44 S L11, L12, L16, L17
E PHYLOGENY/CT
E E3+ALL/CT
L19 46381 S E4+NT/CT
L20 11 S L19 AND L6
L21 11 S L19 AND L17
L22 48 S L21, L19
E GENOME/CT
E E3+ALL/CT
L23 18380 S E6+NT/CT
L24 3 S L6 AND L23
L25 43 S L21, L24
L26 46 S L25 AND D14..,CT
L27 3 S L27 NOT L14..,CT
L28 11 S L17 NOT D14..,CT
L29 11 S L28 AND L11, L14, L15, L19, L23
E ALGORITHM/CT
E E4+ALL/CT
L30 23513 S E12+NT/CT
E DATABASE/CT
L31 13061 S E3+NT/CT
E PROTEIN CONFORMATION+ALL/CT
L32 95913 S E9+NT/CT
E PROTEIN FOLDING/CT
E E3+ALL/CT
L33 8913 S E4+NT/CT
L34 163 S L6 AND L31-L34
L35 33 S L35 AND L25-L30
L36 11 S L36-L30 NOT L16
L37 7 S L37 NOT D14..,CT
L38 49 S L36, L38

FILE 'MEDLINE' ENTERED AT 12:47:05 ON 12 FEB 2001
SEL L34 AN 7 17-15 19 29 30 33-38 40 41

L40 15 S E1-E15
L41 36 S L39 NOT L40
L42 11 S L41 AND L7
L43 36 S L41, L42
L44 17 S L43 AND CALIGRI
L45 36 S L45, L44

FILE 'MEDLINE' ENTERED AT 12:57:13 ON 12 FEB 2001

FILE 'BIOVIS' ENTERED AT 12:57:57 ON 12 FEB 2001

E MARCOTTE R/AU
L46 32 S E3, E6, E7
E EISENBERG D/AU
L47 24 S E3-E4, E13-E16
E THOMPSON M/AU
L48 611 S E5, E21, E22
E THOMPSON MICH/AU
L49 44 S E4, E11
E PELLEGRINI M/AU
L50 203 S E3-E12, E25
E YEATES T/AU

L51 48 S E3-EG
 L52 1199 S L46-L51
 L53 1018263 S (*10064 OR *10054)/CC
 L54 1281 S L53 AND L52
 L55 49 S 04-60/CC AND L54
 L56 214 S 10366/CC AND L54
 L57 8 S 01-66/CC AND L54
 L58 42 S L57, L57
 L59 10 S L56 AND L58
 L60 401 S L52 AND 101520, CC OR CONFERENCE/DT
 L61 454 S L52 AND CONFERENCE OR CONGRESS OR POSTER OR SYMPOS? OR MEETI
 L62 455 S L52, L51
 L63 51 S L61 NOT CONFERENCE/DT
 L64 10 S L53 NOT ARTICLE/DP
 SEL IN 10 S 4
 L65 1 S E1-E3
 L66 11 S L64 NOT L65
 L67 434 S L66, L66
 L68 62 S L67 AND L68
 L69 63 S L68 AND L68
 L70 15 S L68 AND L68
 L71 5 S L68 AND L68
 L72 10 S L68-L71
 L73 10 S L71 AND LIVER OR REPLACEMENT OR MELLITIN OR MILK OR ENCEPHAL
 L74 11 S L71 NOT L7
 L75 12 S L71 NOT L67-L74
 L76 13 S L71 AND COCOGNT OR EVOLUTION? OR PATTERN? OR ALGORITHM? OR PR
 L77 13 S L76 NOT L71
 L78 17 S L71, L77

FILE 'BIOCTS' ENTERED AT 13:19:50 ON 12 FEB 2001

FILE 'WPIX' ENTERED AT 13:26:37 ON 12 FEB 2001
E MARCOTTE E/AU

L79 1 S E3
 E EISENBERG D/AU
 E THOMPSON M/AU
 L80 159 S E?, E13, E14
 E PELLEGRINI M/AU
 L81 12 S E3-E6
 E YEATES T/AU
 L82 6 S E3, E4
 E EISENBERG D/AU
 L83 13 S E3-E6
 L84 136 S L78-L83
 L85 13 S L84 AND G06F/IC
 L86 7 S L84 AND T01-J7/MC
 L87 13 S L85, L86
 L88 3 S L87 AND (PROTEIN OR PEPTIDE OR POLYPEPTIDE)
 L89 14 S L84 AND C07K/IC
 L90 10 S L84 AND (B04-B04A? OR C04-B04A? OR B04-N04? OR C04-N04?) /MC
 L91 3 S L84 AND V93/M0, M1, M2, M3, M4, M5, M6
 L92 20 S L89-L91
 L93 11 S L87 AND L82
 L94 5 S L88, L83
 L95 12 S L81 NOT L84

FILE 'WPIX' ENTERED AT 13:32:46 ON 12 FEB 2001
E G06F016/IC

L96 114-2 S E7A
 L97 414253 S (T01 OR T04)/DC
 L98 309436 S G06F/IC
 L99 1154 S T01-J74A/MC
 E T01-S-ALL/MC
 L100 6953 S T01-S1/MC
 L101 447966 S L96-L100

L112 464 S L101 AND (PROTEIN OR PEPTIDE OR POLYPEPTIDE OR C07K/IC OR (B0
 L113 104 S L102 AND FUNCTION?
 L114 16 S L103 AND PROFILE
 L115 4 S L103 AND (N HOMOLOG? OR NON HOMOLOG?)
 L116 56 S L104 AND MULTIPLET?
 L117 16 S L105 AND L104-L106
 L118 4 S L106 AND EVO/L?
 L119 31 S L107 AND GENOM.
 L119 34 S L107 AND PALIGN?
 L119 23 S L107 AND L108-L110
 L119 34 S L107, L111
 L119 50 S L107, L115, L116, L117
 L119 51 S L107 AND L110
 L119 52 S L107, L114
 L119 53 S L107 AND (P01-K04 OR C12-K04 OR D05-H09)/MC
 L119 54 S L107 AND (P01-K04? OR C12-K04?)/MC
 L119 55 S L107, L117 AND L108
 L119 56 S L118, L115 NOT L107-L09
 L119 57 S L118 AND GOF/ICM
 L119 58 S L118 NOT L106
 L119 47 S L119 AND GOF/ICM
 L120 7 S L120, L122
 L120 10 S L120 NOT L107

FILE 'WPIX' ENTERED AT 14:36:17 ON 12 FEB 2001

L121 6 S L101 AND PHYLOG.
 L126 17 S L101 AND PHYLOG?
 L127 18 S L116 NOT L06-L09

FILE 'MEDLINE' ENTERED AT 14:41:57 ON 12 FEB 2001

L131 460416 S L7, L111, L134 OR HEADING #AMES+NT/CT OR MOLECULAR SEQUENCE DATA
 L132 2337 S L127 AND L129
 L132 110242 S L127 AND L23
 L133 47028 S L128 AND L23, L24
 E PROTEIN STRUCTURE/CT
 E E4+ALL/CT
 L134 13561 S PHYLOG AND L117-L131
 L135 12536 S PHYLOGNY+NT/CT AND L129-L131
 L136 13561 S L132, L133
 L137 6774 S L134 AND ?HOMOLOG?
 L138 2554 S L134 AND FUNCTION?
 L139 1637 S L134 AND L136
 L140 1115 S L137 AND L137/CT
 L140 1184 S L137 AND D137/CT
 L140 70 S L139 AND L139
 L141 1182 S L139 NOT L39-L49
 L142 19 S L141 AND PROTEIN FOLD?
 L143 802 S L139 AND PY =1309
 L144 7 S L143 AND L141
 L145 46 S L143 AND L144
 L146 92 S L144, L145
 L147 46013 S SEQUENCE HOMOLOGY, AMINO ACID/CT
 L148 1612 S EVOLUTION, MOLECULAR/CT AND L147
 L149 942 S L147 AND PY =1309
 L150 15 S L111 AND L144
 L151 373 S L7 AND L144
 L152 36 S L156 AND L151
 L153 29 S L23 AND L150, L151
 L154 372 S L154, L151 AND D127/CT NOT L153
 L155 14 S L154 NOT AB/FA
 L156 397 S L154 NOT L153
 L157 23 S L156 AND L117/CT

FILE 'MEDLINE' ENTERED AT 14:57:21 ON 12 FEB 2001

L158 56 S L156 AND L37, L34
 L159 156 S L156 AND PHYLOG?

L160 20 S L159 AND L158
L161 56 S L158, L160
SEL L161 AN 5 11 15 19 24 37 40 54
L162 8 S E1-E8